PB85-234581

Mathematical and Empirical Receptor Models Workshop: Quail Roost 2 Reports

(U.S.) Environmental Protection Agency
Research Triangle Park, NC

1984

U.S. Department of Commerce
National Technical Information Service
NTIS

PB85-234581

EPA/600/J-84/317
JOURNAL ARTICLE

MATHEMATICAL AND EMPIRICAL RECEPTOR MODELS WORKSHOP

QUAIL ROOST II REPORTS

Atmospheric Environment, 18(8):1499-1582, August 1984

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA/600/J-84/317	2.	3. RECIPIENT'S ACCESSION NO. PB85 234581 /AS
4. TITLE AND SUBTITLE **MATHEMATICAL AND EMPIRICAL RECEPTOR MODELS WORKSHOP QUAIL ROOST II REPORTS** (Journal Version)		5. REPORT DATE 1984
		6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) R. K. Stevens, Editor		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Atmospheric Sciences Research Laboratory Office of Research and Development U.S. Environmental Protection Agency Research Triangle Park, N. C. 27711		10. PROGRAM ELEMENT NO. CDTAID-01 2108 (FY-83)
		11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS Atmospheric Sciences Research Laboratory Office of Research and Development U.S. Environmental Protection Agency Research Triangle Park, N. C. 27711		13. TYPE OF REPORT AND PERIOD COVERED
		14. SPONSORING AGENCY CODE EPA/600/09

15. SUPPLEMENTARY NOTES

Atmospheric Environment, 18(8):1499-1582, August 1984

16. ABSTRACT

On March 14-17, 1982, the U.S. Environmental Protection Agency (EPA) sponsored the Mathematical and Empirical Receptor Models Workshop (Quail Roost II) at the Quail Roost Conference Center in Rougemont, NC. EPA invited 35 scientists to participate. Results of receptor modeling applied to real and simulated data sets were discussed, evaluated, and structured in a manner that permitted intercomparisons of the models. The participants also were asked to prepare reports on the state of the art of receptor modeling based on the results of their analyses of the simulated and real data. Topics covered in the reports are as follows: Overview of the Mathematical and Empirical Receptor Models Workshop; Review of Receptor Model Fundamentals; Interlaboratory Comparison of Source Apportionment Procedures: Results for Simulated Data Sets; Chemical and Physical Analyses of Houston Aerosol for Interlaboratory Comparison of Source Apportionment Procedures; Interlaboratory Comparison of Receptor Model Results for Houston Aerosol; and Considerations for Design of Source Apportionment Studies.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

18. DISTRIBUTION STATEMENT RELEASE TO PUBLIC	19. SECURITY CLASS *(This Report)* UNCLASSIFIED	21. NO. OF PAGES 86
	20. SECURITY CLASS *(This page)* UNCLASSIFIED	22. PRICE

EPA Form 2220-1 (Rev. 4-77) PREVIOUS EDITION IS OBSOLETE

ATTENTION

AS NOTED IN THE NTIS ANNOUNCEMENT,
PORTIONS OF THIS REPORT ARE NOT LEGIBLE.
HOWEVER, IT IS THE BEST REPRODUCTION
AVAILABLE FROM THE COPY SENT TO NTIS.

NOTICE

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

This article documents research that was conducted under the auspices of the U.S. Environmental Protection Agency (USEPA). As an arm of the government, the USEPA has a paid up, non-exclusive, irrevocable, worldwide license to reproduce, prepare derivative works, and distribute to the public all of the material contained herein, and to grant others acting on its behalf the right to do so. The USEPA grants to the National Technical Information Service, Department of Commerce, the right to act as its agent in the distribution of this material.

Atmospheric Environment Vol. 18, No. 8, pp. 1499–1506, 1984
Printed in Great Britain.

Pergamon Press Ltd

OVERVIEW OF THE MATHEMATICAL AND EMPIRICAL RECEPTOR MODELS WORKSHOP (QUAIL ROOST II)

ROBERT K. STEVENS and THOMPSON G. PACE
U.S. Environmental Protection Agency Research Triangle Park, NC 27711, U.S.A.

(*First received 9 June* 1983 *and in final form* 19 *January* 1984)

Abstract—On 14–17 March 1982, the U.S. Environmental Protection Agency sponsored the Mathematical and Empirical Receptor Models Workshop (Quail Roost II) at the Quail Roost Conference Center, Rougemont, NC. Thirty-five scientists were invited to participate. The objective of the workshop was to document and compare results of source apportionment analyses of simulated and real aerosol data sets. The simulated data set was developed by scientists from the National Bureau of Standards. It consisted of elemental mass data generated using a dispersion model that simulated transport of aerosols from a variety of sources to a receptor site. The real data set contained the mass, elemental, and ionic species concentrations of samples obtained in 18 consecutive 12-h sampling periods in Houston, TX. Some participants performed additional analyses of the Houston filters by X-ray powder diffraction, scanning electron microscopy, or light microscopy. Ten groups analyzed these data sets using a variety of modeling procedures. The results of the modeling exercises were evaluated and structured in a manner that permitted model intercomparisons. The major conclusions and recommendations derived from the intercomparisons were: (1) using aerosol elemental composition data, receptor models can resolve *major* emission sources, but additional analyses (including light microscopy and X-ray diffraction) significantly increase the number of sources that can be resolved; (2) simulated data sets that contain up to 6 dissimilar emission sources need to be generated, so that different receptor models can be adequately compared; (3) source apportionment methods need to be modified to incorporate a means of apportioning such aerosol species as sulfate and nitrate formed from SO$_2$ and NO$_x$ respectively, because current models tend to resolve particles into chemical species rather than to deduce their sources and (4) a source signature library may be required to be compiled for each airshed in order to improve the resolving capabilities of receptor models.

INTRODUCTION

On 14–17 March 1982, the U.S. Environmental Protection Agency (EPA) sponsored the Mathematical and Empirical Receptor Models Workshop (Quail Roost II) at the Quail Roost Conference Center in Rougemont, NC. EPA invited 35 scientists to participate. Results of receptor modeling applied to real and simulated data sets were discussed, evaluated and structured in a manner that permitted intercomparisons of the models. The participants also were asked to prepare reports on the state of the art of receptor modeling based on the results of their analyses of the simulated and real data.

The predecessor of Quail Roost II was a March 1980 workshop also held at the Quail Roost Conference Center. At that 2½-day meeting (the Quail Roost I Receptor Modeling Workshop), scientists met to discuss the status of mathematical and analytical chemical procedures to determine the natural and man-made sources that contribute to atmospheric particles. The objective was to obtain data and information for a report on the state-of-the-art of receptor modeling (sometimes referred to as 'source apportionment modeling'). This report was to contain recommendations for research projects that would advance knowledge in this relatively new area of air pollution science.

The EPA report that resulted from Quail Roost I (Watson *et al.*, 1981) presented three major recommendations:

(1) Develop a *simulated* air quality and source contribution data set perturbed by random error to test the predictions of several different types of receptor models against known source contributions.

(2) Obtain a *real* air quality data set to test the consistency of receptor models and source apportionment procedures currently used by various groups in the United States.

(3) Standardize validated models so that they can be readily implemented by Federal, state, and local agencies.

Following Quail Roost I, several EPA and National Bureau of Standards (NBS) scientists worked to generate simulated and real air quality data sets consistent with recommendations 1 and 2. Three simulated data sets were developed. Each consisted of lists of mass and elemental concentrations for 40 sequential fine-fraction aerosol observations at a single receptor site, and elemental composition profiles for up to 13 different source types from which these observations were generated.

The real data set was derived from EPA measurements in Houston, TX, during September 1980. The data set was based on 18 sequential fine- and coarse-fraction observations at a single field site, and included mass, elemental, and ionic species concentrations from beta gauge, X-ray fluorescence (XRF), instrumental neutron activation analysis (INAA), and wet chemical measurements performed on the collection filters. Since source profiles specific to Houston's major

Reproduced from
best available copy.

aerosol emission sources were unavailable, source profiles obtained in other cities were made available along with the ambient data set. In addition, certain participants analyzed filter samples by X-ray powder diffraction (XRD), light microscopy (LM) and scanning electron microscopy with automated image analysis and X-ray energy spectroscopy (SAX). Johnson *et al.* (1984) describe all of the Houston sampling and analytical procedures in more detail.

The availability of these data sets provided a basis for a second Quail Roost receptor model workshop. The principal objective of this Quail Roost II workshop was to document the current status of source apportionment methodology by determining, for the simulated data set, how well the different receptor models were able to *reproduce* the true source contributions, and by comparing, for the real (Houston) data set, the *consistency* of results from different source apportionment methods when the true source contributions were unknown. It was anticipated that these exercises would lead to a better understanding of the strengths and weaknesses of specific receptor models, and to recommendations for improving their performance.

The simulated and real data sets and associated source profiles were distributed in advance of Quail Roost II on nine-track magnetic tape and in hard copy form. Between the time of receipt and the beginning of the workshop (five weeks for the real data set, but less than two weeks for the simulated sets), each participating laboratory was expected to apply its particular source apportionment method to either or both data sets and prepare a report that would contain the following information:

(1) method(s) of source apportionment, including any special means for making judgments;

(2) estimates of the numbers and identities of source types;

(3) for each source thought to be present, estimates of its contributions ($\mu g\ m^{-3}$) to the individual periods and

(4) uncertainty estimates for each component, together with limits for components believed to be present but not detectable.

Table 1 lists the participating organizations that submitted results of source apportionment treatments of the simulated and real data sets. The University of Illinois at Urbana–Champaign and the Texas Air Control Board submitted multiple linear regression results for both the simulated and real sets. Environmental Research and Technology, Inc., Radian Corporation, and NEA, Inc. submitted results obtained by unified factor analysis-chemical mass balance (CMB) models for one of the simulated data sets. Other contributions to the analysis are listed in Table 1.

At the workshop, the participants were asked to prepare a summary of mathematical procedures and a hierarchy of source apportionment strategies that would assist state and local agencies in designing receptor model projects. The mathematical procedures are summarized by Henry *et al.* (1984), and the source apportionment strategies are presented by Gordon *et al.* (1984).

The intercomparisons of receptor models by Currie *et al.* (1984) and Dzubay *et al.* (1984) demonstrate the need to establish a standardized format for reporting source apportionment results, and highlight the need for source signatures to include not only elemental profiles but also XRD, scanning electron microscopy (SEM), isotopic (^{14}C), and LM characteristics of both fine- and coarse-fraction emissions.

The participants decided that a glossary of source apportionment modelling terms should be generated and circulated among receptor modelers for comment and revision. Such a glossary was compiled and is included at the end of this report; it contains those terms most commonly used in source apportionment

Table 1. Participants who submitted results for the simulated and real air quality data sets

Participant	Method	Data set Simulated	Real
Environmental Research and Technology, Inc.	Effective variance CMB	X	X
NEA, Inc.	Effective variance CMB	X	X
U.S. Environmental Protection Agency (Environmental Sciences Research Laboratory)	Effective variance CMB/factor analysis	X	
Radian Corporation	Weighted least-squares CMB (real set) and weighted ridge regression CMB (both sets)	X	X
Texas Air Control Board	Multiple linear regression by tracers/factor analysis	X	X
South Dakota School of Mines and Technology	X-ray powder diffraction/effective variance CMB		X
State University of New York College of Environmental Science and Forestry	Q-mode factor analysis (Full *et al.*, 1981) of SAX data		X
Microlab Northwest	Light microscopy of coarse fraction		X
Lawrence Berkeley Laboratory	X-ray powder diffraction		X
Harvard University	Weighted CMB/factor analysis	X	
University of Illinois at Urbana-Champaign	Target transformation factor analysis	X	X

and receptor modeling. We have tried, of course, to provide clear and unambiguous definitions, but we recognize that the language, as well as the method, is in a state of evolution.

DISCUSSION OF RESULTS

Simulated data set

NBS sent to the participants three simulated data sets for receptor modeling analysis. However, only Sets I and II were analyzed by the participants, since they had only two weeks prior to the workshop to prepare their source apportionment reports.

Data Set I consisted of mass and elemental concentrations for 40 sequential fine-fraction aerosol observations, along with source signatures for 13 different sources that potentially could have impacted the receptor site (only 8 of these were in fact active). Unknown to the participants, an additional unspecified source also contributed to the Set I observations. Data Set II consisted of 40 observations, with 13 known sources contributing to the receptor site. Data Sets I and II also contained at least four groups of sources within which signatures were similar. Data Sets I and II were designed to determine whether any of the receptor mathematical treatments could resolve sources emitting particles of similar elemental composition, and whether any of the models could detect and calculate the mass contributed to the receptor site by an unknown source.

None of the conventional CMB methods detected the unknown source, but two of the factor analysis methods and a multiple linear regression model were able to deduce the mass contributed by the unknown source. All of the receptor models had difficulty resolving sources with similar signatures. These difficulties created substantial uncertainties in the apportionment of those sources. In a companion paper, Currie *et al.* (1984) describe the features and development of the simulated data sets and intercompare the receptor model results obtained for those sets.

Real (Houston) data set

The real air quality data provided to the Quail Roost II participants were derived from an EPA visibility study conducted in Houston during the summer of 1980. Although we collected the aerosols on three different filter media in order to perform a wide range of chemical analyses, our sampling protocol was not specifically designed for receptor modeling. For example, particle loadings on the Houston filters were about a factor of 2 less (in terms of average mass per filter surface area) than optimum for XRD analysis. Also, we did *not* make a special effort to collect and preserve the ambient aerosol nitrates. Shaw *et al.* (1982) have shown that ambient nitrates can be reliably measured through use of a denuder difference sampling system. However, if particle sampling is conducted *without* these precautions (as it was in Houston), nitrate concentrations can be underestimated by as much as a factor of 10 (Stevens, 1979).

Figure 1 shows that the conventional receptor or factor analysis models could only apportion the fine and coarse fractions to three chemical categories (carbon, nitrate and sulfate) and four general sources (vehicle exhaust, steel, marine and crustal). On the other hand, LM and SAX were able to resolve 18 sources. Even though the sample mass loadings were inadequate for optimum application of XRD, participants from Lawrence Berkeley Laboratory and the South Dakota School of Mines and Technology were able to determine that most of the calcium in the coarse particles was of local soil origin based on the identification of a form of $CaCO_3$ indigenous to Houston.

Considerations for design of source apportionment studies

A group of participants examined and evaluated the preliminary results presented at Quail Roost II and prepared a report (Gordon *et al.*, 1984) that contains recommendations for the design of source apportionment studies by private or governmental institutions. One of the key issues addressed by this workshop subgroup was the development of protocols for data collection and analysis appropriate to different levels of source apportionment. The levels of effort described below assume that the goal of the source apportionment study is to assist development of control strategies that will reduce particulate mass concentrations so that the airshed will be in compliance with the National Ambient Air Quality Standard for particulate matter. The elements (and/or chemical species) that would be measured as input to the receptor model would vary from airshed to airshed; Table 2 lists some species that are typically determined.

The Quail Roost II participants recommended, in general, that source apportionment studies be divided into three levels of intensity based principally on resource availability:

Level I: Minimal resources. This effort uses existing source signature libraries, existing ambient air quality data (mass and elemental composition of fine and coarse particles), local emissions inventories and literature values for soil composition. These data suffice for an initial apportionment by CMB and for factor analysis to determine if there are contributing sources that are not included in the emissions inventories. The cost for this level of effort is between $10,000 and $50,000, depending on the size of the airshed and on whether additional sample analyses are required. Most of the effort is spent in gathering emissions data and inputting data to the receptor model.

Level II: Moderate resources. This level of effort includes all of the Level I work plus acquisition of source elemental profiles through the grab sampling of fugitive sources, soil, and major primary particle emitters in the airshed. The grab samples are resuspended and recollected in the same manner and on the same filter material used to collect particles at the

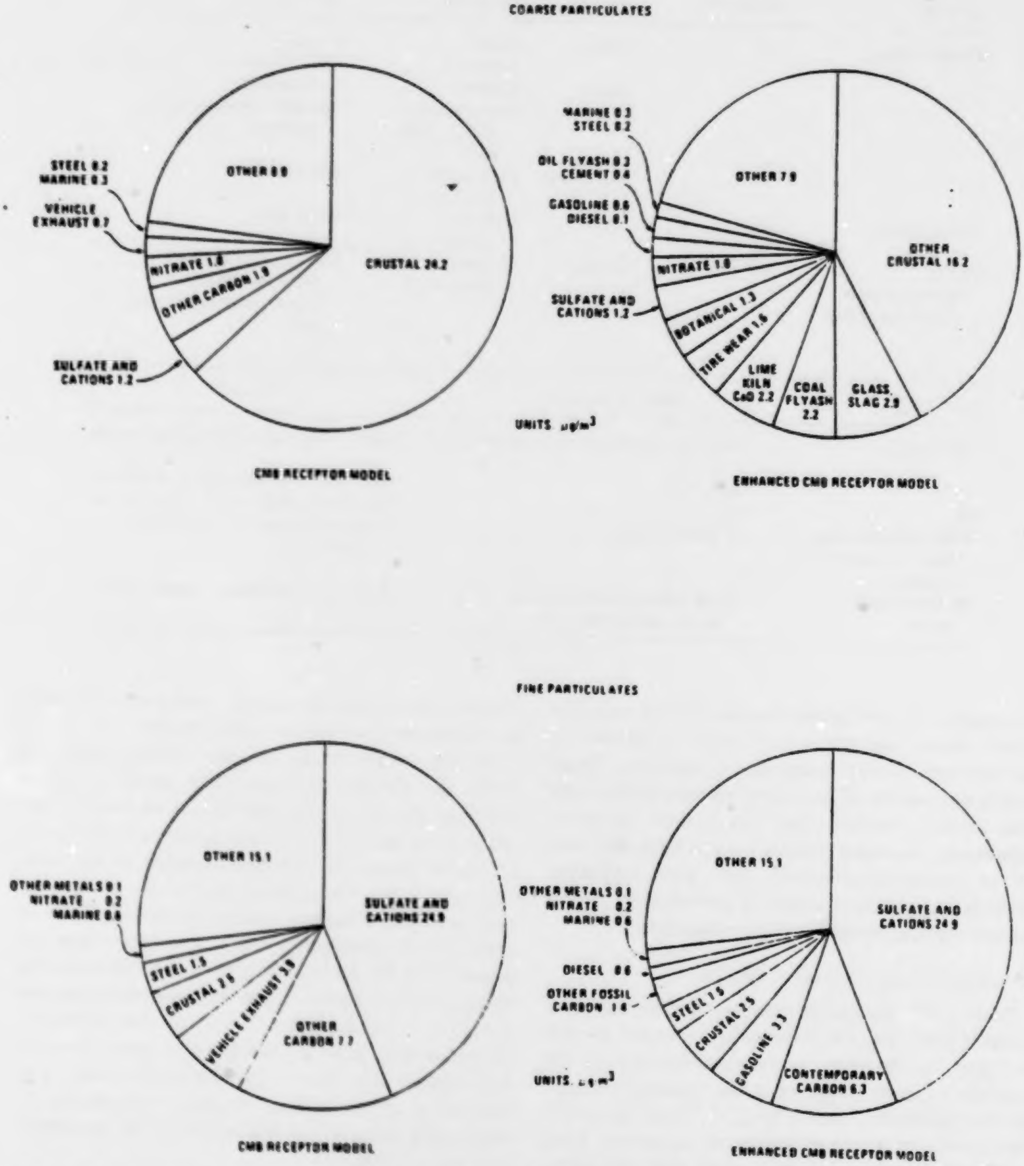

Fig. 1. Basic and enhanced chemical mass balance (CMB) results for fine and coarse particles collected in Houston, TX, 10 September 1980, 0700–1900 CDT. Enhanced results were obtained through the use of light microscopy, scanning electron microscopy with automated image analysis and X-ray energy spectroscopy, and X-ray powder diffraction.

receptor site(s). Most of the Level II expenditures are in the collection, redeposition, and analysis of grab samples. The Level II cost ranges from $25,000 to $100,000, depending on the size of the airshed under study.

Level III: Maximum resources. This level of effort calls for (1) collection of aerosols at receptor sites where maximum concentrations may occur, (2) sampling of emissions from major sources with a dilution sampler that provides a sample suitable for chemical analysis, (3) analysis of filters from the sources and receptor sites by XRF and INAA and (4) use of XRD, SEM and LM to resolve sources with similar emission profiles. This level of effort costs between $50,000 and $500,000, again depending on the size of the airshed under investigation. Most of the costs are in the collection and analysis of source and ambient samples.

This workshop subgroup also considered the samplers and other data collection devices that should be part of a complete receptor model station. Table 2 is based on their recommendations. The subgroup further recommended that multivariate and wind trajec-

Table 2. Instrumentation to collect ambient data for receptor modeling

Particulate sampler	Flow rate (ℓ min^{-1})	Filter medium	Species measured	Analytical method(s)
Dichotomous	17	Teflon	Mass Elements	β-gauge XRF, INAA
		Quartz	Carbon Ionic species (SO_4^{2-}, NO_3^-, etc.)	Combustion Ion chromatography/chemistry
		Nuclepore	Individual particles	SEM/XRF
Dichotomous	50	Teflon	Crystals, minerals	XRD. LM
Hi-vol (with cyclone) 2.5 μm cut point	500–1000	Quartz	Organic compounds	Gas chromatography/mass spectrometry liquid chromatography, etc.

Gas sampler	Species measured	Output
Total sulfur	SO_2 and other sulfur-containing gases	Real time, need data-logging system
NO_x	NO, NO_2	Real time, need data-logging system
O_3	O_3	Real time, need data-logging system
Polyurethane, cold trap, or packed columns	Organic vapors	Later extraction and analysis in laboratory
Meteorological tower	Wind speed, wind direction, temperature, dew point	Real-time, need data-logging system

tory receptor models be use wherever there is a need to obtain source signatures not readily available by conventional source sampling procedures. These models are capable of providing source signatures by their unique treatment of air quality data. However, multivariate statistical models require large data sets (40 or more observations), and wind trajectory methods require short sampling periods (6 h or less) and appropriate meteorological measurements.

Recommendations for future research

Table 3 lists the marker elements used by the Quail Roost II participants in their various receptor models to report the emission sources contributing to the particulate matter measured at the Houston receptor site. The sulfate and nitrate listed in Table 3 do not, in the absence of appropriate source signatures, have marker elements that reliably deduce their origins. At present, we can only speculate as to the origin(s) of the sulfate and nitrate. Even analysis of the Houston filters by SAX and LM failed to deduce their origin. In most airsheds in the United States, sulfate and nitrate and their associated cations represent a substantial portion of the mass of particles below 10 μm in diameter. If source apportionment methodology is to adequately address urban and rural particulate concentration problems, modeling procedures must be developed to determine the origins of the sulfate and nitrate that impact a given receptor site.

Listed below are additional major recommendations based on the results of the model intercomparisons and on the participants' past experience with air quality source apportionment studies:

(1) The comparability of the various computer codes for effective variance CMB needs to be established through testing on a simulated data set that contains no more than six components and for which all source profiles are known. Accuracy of the computer programs would be established by comparing their results with the true source contributions used to construct the simulated sample. The error estimates produced by the various computer programs would be compared. Comparability among computer programs should be established on such a simple data set before further comparisons are made using more complex and difficult sets. Also, a single computer code representative of those used in the study should be thoroughly documented and evaluated for sensitivity to input assumptions.

(2) The categories for reporting results from receptor modeling need to be standardized. The source profiles should be given (or referenced to the literature) along with the source intensities derived from them. One of the difficulties in intercomparing the Quail Roost II results was the ambiguity of some of the source categories used by different participants in their analyses of the Houston data set.

(3) Source signatures for receptor modeling should be obtained in a receptor-compatible manner. That is, the source sample must be collected in such a way that the properties of the sample at the source and at the receptor are as similar as possible, and the analytical

Table 3. Source categories and marker elements used in source apportionment of Houston aerosol

Marker elements	General source category	Specific source categories
Na, Cl	Marine	Marine
Al, Si, K, Ca, Mn, Fe	Crustal	Wind erosion of soil Paved and unpaved roads Construction Limestone crushing and handling Cement Lime kiln Fly ash Slag
C, Br, Pb	Transportation	Noncatalyst vehicles Catalyst cars Diesels
Mn, Fe	Steel	Iron and steel production Steel finishing and handling
Cu, Zn, As, Sm, Sb	Nonferrous metals	
NH_4^+, SO_4^{2-}	Sulfate	Primary emissions in Houston airshed Conversion of SO_2 in Houston airshed Regional background
C	Carbon	Refineries Botanical Vegetative burning Tire wear
NO_3^-	Nitrate	Photochemical

techniques used on the source and receptor samples must be essentially identical. These criteria may be met by dilution sampling of source emissions in which the aerosol is collected on filter media and with collection devices equivalent to those used at the receptor site. This technique also improves the ability of receptor models to deal with condensable particular matter.

(4) The uncertainties associated with source contribution estimates need to be better quantified. At present, different receptor models calculate error estimates in nonequivalent ways.

(5) A future intercomparison should be held in which the various source apportionment methods are applied to air quality and source signature data obtained by EPA in 1982 for the Philadelphia airshed. The Philadelphia source signature and air quality data were obtained by sampling and analysis procedures recommended in this paper. Participants in the intercomparison should be provided with the analytical data at least nine months prior to the workshop to allow sufficient time for analysis and report preparation.

(6) The results of receptor and dispersion modeling should be compared for an urban case in which air quality and meteorological data appropriate to support both models are obtained. The purpose of the comparisons would be to identify the features of each model that can be validated and used to enhance the source apportionment predictions of the other.

CONCLUSIONS

The Quail Roost II workshop exercises led to a number of conclusions that enhance our understanding of source apportionment methodologies. Some of the most important are as follows:

(1) The simulated data set exercises indicated that neither a multivariate nor CMB approach to source apportionment can be considered the method of choice for all circumstances. While a CMB approach is mathematically easier to implement and does not require a large number of samples, it requires reliable characterization (profiles and their variabilities) of all major potential sources for the region to which it is applied. On the other hand, while multivariate analysis does not require comprehensive advance information on source signatures, it does require a very large number of ambient samples. As with CMB, the source resolution capability depends on the specific data base (amount of collinearity, measurement uncertainties, etc.). With the Quail Roost II simulated data sets, only four to seven classes of source could be resolved. This degree of resolution can be sufficient for regulatory purposes when several broad source categories are largely responsible for elevated particulate concentrations.

(2) The three CMB methods investigated gave similar and generally satisfactory results for the simulated data set for which all 13 source profiles were

Reproduced from best available copy.

supplied. The problem of multicollinearity among certain source profiles did, however, seriously degrade the precision of intensity estimates for these sources. This demonstrates the need for additional species characterization and for development of methods to resolve sources with similar elemental profiles.

(3) Multivariate procedures isolated up to seven of the nine sources active in simulated Data Set I. At least one multivariate technique produced a good estimate for the profile of the 'unknown' component. For Set II, target transformation factor analysis (TTFA) isolated only 4 of the 13 sources.

(4) A major outcome of the simulation exercise was a recognition of the need for improved treatment of uncertainties. Even participants employing the same basic model gave widely different uncertainty estimates. Also, there appeared to be no satisfactory treatment of systematic errors in source profiles.

(5) For the Houston data set, CMB and multivariate models were in agreement within a range of 30–40% for the crustal matter source category. However, results for the transportation category obtained by multivariate analysis were about a factor of 3 larger than results obtained by CMB for a mixed population of vehicles. The latter were determined mainly by the abundance of lead in the source signature for vehicle exhaust. The signature used was obtained about two years before the Houston study, and did not take into account the factor-of-2 decrease in the lead content of vehicle exhaust that occurred between 1978 and 1980. In addition, the results by multivariate analysis may include secondary aerosol arising from atmospheric reactions of automotive exhaust. Formation of secondary aerosol was not included in the signatures used with the CMB models. TTFA results indicated that the fine fraction of the transportation category contained less than 15% crustal matter. We conclude that transportation source signatures derived by TTFA in this study are potentially more realistic than those for weighted populations of vehicles derived by means of emissions inventories.

(6) Ridge regression and weighted least-squares results were compared for the case of 18 source profiles. As expected, the resulting estimates of uncertainties for each component were far smaller with the ridge regression method than with the least-squares method. Nonetheless, the effects of multicollinearity were so severe that of the 18 source contributions estimated using ridge regression, only two values were significantly different (statistically) from 0 (i.e. concentration $> 2\sigma$). Therefore, in resolving sources for this data set, the ridge regression model offered little mathematical advantage over conventional least-squares.

(7) The unavailability of source profiles for the major emission sources in Houston and the lack of representative soil samples limited the accuracy of the CMB results. While the number of sampling periods (18) was adequate for CMB, the number was insufficient for accurate multivariate analysis, which requires at least 40 samples.

(8) LM and SAX successfully resolved the general categories of the Houston samples into several specific source related categories. The combining results from these methods with results from mathematical receptor models that use bulk chemical composition, the most comprehensive source apportionment results can be obtained.

(9) XRD may offer a viable alternative to other analytical techniques for identifying certain sources of crustal matter. In this study, even though the full potential of XRD was not demonstrated because the aerosol sampling procedure did not provide optimum deposit thicknesses, the XRD analysis did give quantitative results for calcite, quartz, mascagnite and dolomite in the coarse fraction.

Acknowledgements—The authors thank Northrop Services, Inc., for its help on this project. Specific thanks go to Leslie Silkworth, who coordinated the Quail Roost II workshop, Slater Raymond, who edited the reports prepared by workshop attendees, and Connie van Oosten, who typed the manuscripts. Also, the authors express their gratitude to the participants themselves, who contributed so freely of their time and expertise in preparing the various receptor modeling reports.

REFERENCES

Currie L. A., Gerlach R. W., Lewis C. W., Balfour W. D., Cooper J. A., Dattner S. L., De Cesar R. T., Gordon G. E., Heisler S. L., Hopke P. K., Shah J. J., Thurston G. D. and Williamson H. J. (1984) Interlaboratory comparison of source apportionment procedures: results for simulated data sets. *Atmospheric Environment* **18**, 1517–1537.

Dzubay T. G., Stevens R. K., Balfour W. D., Williamson H. J., Cooper J. A., Core J. E., De Cesar R. T., Crutcher E. R., Dattner S. L., Davis B. L., Heisler S. L., Shah J. J., Hopke P. K. and Johnson D. L. (1984) Interlaboratory comparison of receptor model results for Houston aerosol. *Atmospheric Environment* **18**, 1555–1566.

Full W. E., Ehrlich R. and Klovan J. E. (1981) Extended Q model—Objective definition of external end members in the analysis of mixtures. *Math. Geol.* **13**, 331–334.

Gordon G. E., Pierson W. R., Daisey J. M., Lioy P. J., Cooper J. A., Watson J. G., Jr. and Cass G. R. (1984) Considerations for design of source apportionment studies. *Atmospheric Environment* **18**, 1567–1582.

Henry R. C., Lewis C. W., Hopke P. K. and Williamson H. J. (1984) Review of receptor model fundamentals. *Atmospheric Environment* **18**, 1507–1515.

Johnson D. L., Davis B. L., Dzubay T. G., Hasan H., Crutcher E. R., Courtney W. J., Jaklevic J. M., Thompson A. C. and Hopke P. K. (1984) Chemical and physical analyses of Houston aerosol for interlaboratory comparison of source apportionment procedures. *Atmospheric Environment* **18**, 1539–1553.

Shaw R. W., Stevens R. K., Bowermaster J., Tesch J. W. and Tew E. (1982) Measurements of atmospheric nitrate and nitric acid: the denuder difference experiment. *Atmospheric Environment* **16**, 845–859.

Stevens R. K. (1979) Current methods to measure atmospheric nitric acid and nitrate artifacts. EPA-600/2-79-051, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Watson J. G. *et al.* (1981) Receptor models relating ambient suspended particulate matter to sources. EPA-600/2-81-

039, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

GLOSSARY

Aerosol: a suspension of solid or liquid particles in a gas.

Backward stepwise multiple regression: similar to ordinary stepwise regression, but with the independent variables successively *removed* according to statistical criteria.

Calcite: a crystalline form of $CaCO_3$.

Cenosphere: hollow, spherical, glassy particles characteristic of coal-fired power plants.

Chemical element balance (CEB): an older term that is synonymous with chemical mass balance.

Chemical mass balance (CMB): a mathematical procedure for source apportionment in which the chemical species measured at the receptor are matched, by linear regression techniques, to those calculated from a group of suspected sources of known chemical composition.

Chemical species balance (CSB): synonymous with chemical mass balance.

Clay minerals: a class of layered SiAl-rich minerals often found as small particles in soils.

Coefficient of fractionation: the fractional change in particle composition that occurs between source and receptor.

Communality: the fraction of total variance of an observable feature explained by factors derived from factor analysis of the data set.

Composite receptor method: the combined application of two or more receptor models to a common set of ambient air quality data.

Dispersion model: synonymous with source model.

Effective variance weighted least-squares: a statistical technique used in chemical mass balance that weights the solution in accordance with uncertainties in the ambient data and source profiles.

Emission factor: the rate of pollutant emission expressed in relation to effluent volume ($g\,SCF^{-1}$), distance ($g\,mi^{-1}$), or other units representing source activity.

Emission inventory: the mass of pollutant emitted from a source over a specified period of time (usually 1 y) calculated on the basis of an emission factor, units of material processed per time interval, and operating schedule.

Emission rate: mass of a pollutant emitted per unit time.

Emissions inventory: a tabulation of emission inventories from all sources within a geographical region.

Factor analysis: a group of related statistical techniques whose purpose is to represent a set of multivariate data in terms of a smaller number of hypothetical variables, i.e. factors. When applied to ambient aerosol measurements, the factors can be related to both source emissions and to atmospheric chemistry and physics through appropriate selection of variables. Factor analysis includes principal component analysis, classical factor analysis, and target transformation factor analysis.

Feldspars: a group of silicate minerals; principally, aluminosilicates of potassium, sodium and calcium.

Hematite: a nonmagnetic crystalline form of Fe_2O_3.

Hybrid model: the use of source emission rates and/or dispersion information in a receptor model.

Magnetite: a magnetic crystalline form of Fe_2O_3.

Marker element: synonymous with tracer element.

Mascagnite: a crystalline form of $(NH_4)_2SO_4$.

Multiple regression analysis: a statistical technique for analyzing the relationship between a dependent variable and a set of independent or predictor variables.

Multicollinearity: a circumstance in which a source signature is nearly the same as a linear combination of one or more other source signatures.

Multivariate analysis (MVA): a collective term that includes all statistical techniques that apportion sources on the basis of the common variability in the chemical species as measured at the receptor.

Pollen: biological particles emitted from trees, plants, and flowers and often found in the coarse size fraction of suspended particulate matter.

Principle component analysis: a version of factor analysis.

Q-mode factor analysis: analysis of the relationship between samples in which the data may be normalized to ensure equal weight.

Quartz: a crystalline form of SiO_2.

Reactive species: a species that can undergo chemical transformation in transport between source and receptor.

Receptor model: a specified mathematical procedure for identifying and quantifying the sources of ambient air contaminants at a receptor primarily on the basis of concentration measurements made at that receptor.

Receptor-oriented model: synonymous with receptor model.

Ridge regression: a statistical technique used in chemical mass balance for dealing with multicollinearity.

R-mode factor analysis: analysis of the relationship between species within a sample in which the data may be normalized to ensure equal weight.

Source apportionment method: a method for apportioning ambient species concentrations to emission sources by means of receptor, dispersion, or hybrid models.

Source model: a specified mathematical procedure for computing the ambient concentration of a species on the basis of emission rates of sources and mathematical representations of meteorological dispersion and chemical reactions in the atmosphere.

Source-oriented model: synonymous with source model.

Source profile: the average chemical composition (e.g. $mg\,g^{-1}$) of the emissions from a source.

Source signature: synonymous with source profile.

Stepwise multiple regression: multiple regression analysis in which independent variables are entered in single steps according to statistical criteria. The variable that explains the greatest amount of variance in the dependent variable is entered first; other variables are then entered stepwise in the order that explains the greatest amount of variance unexplained by variables already in the equation.

Target transformation factor analysis (TTFA): a receptor model approach that combines principle component analysis, target transformation rotation techniques, and mass balance analysis to scale the results into absolute terms.

Target transformation rotation: the mathematical rotation of an eigenvector obtained by a factor analysis toward a defined source profile.

Time series analysis: a set of statistical methods to separate time-dependent from time-independent components in a data set.

Tracer element: a single element used to identify and apportion a specific source or source type.

Tracer model: a version of chemical mass balance that uses only one chemical species per source to calculate the source contribution.

Tracer species: a component (element, chemical ion, color, etc.) unique to a particular source type.

Variance: the square of the standard deviation of a variable.

VARIMAX rotation: an orthogonal rotation scheme used in factor analysis to repartition the variance among the retained factors.

Atmospheric Environment Vol. 18, No. 8, pp. 1507–1515, 1984
Printed in Great Britain.

0004-6981/84 $3.00 + 0.00
© 1984 Pergamon Press Ltd.

REVIEW OF RECEPTOR MODEL FUNDAMENTALS*

RONALD C. HENRY†
Civil Engineering Department, University of Southern California, Los Angeles, CA 90089-0231, U.S.A.

CHARLES W. LEWIS
U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, U.S.A.

PHILIP K. HOPKE
Institute for Environmental Studies, University of Illinois at Urbana-Champaign, Urbana, IL 61801, U.S.A.

and

HUGH J. WILLIAMSON
Radian Corporation, Austin, TX 78766, U.S.A.

(*First received 9 June* 1983 *and in final form* 19 *January* 1984)

Abstract—There are several broad classes of mathematical models used to apportion the aerosol measured at a receptor site to its likely sources. This paper surveys the two types applied in exercises for the Mathematical and Empirical Receptor Models Workshop (Quail Roost II): *chemical mass balance models and multivariate models*. The fundamental principles of each are reviewed. Also considered are the specific models available within each class. These include: tracer element, linear programming, ordinary linear least-squares, effective variance least-squares and ridge regression (all solutions to the chemical mass balance equation), and factor analysis, target transformation factor analysis, multiple linear regression and extended Q-mode factor analysis (all multivariate models). In practical application of chemical mass balance models, a frequent problem is the presence of two or more emission sources whose signatures are very similar. Several techniques to reduce the effects of such multicollinearity are discussed. The propagation of errors for source contribution estimates, another practical concern, also is given special attention.

Key word index: Chemical mass balance, multivariate receptor models, hybrid models, factor analysis, target transformation factor analysis, Q-mode factor analysis, receptor models, multicollinearity, ridge regression model, singular value decomposition, variance inflation factor.

INTRODUCTION

The development of receptor-based models has progressed significantly in the past few years (see Watson *et al.*, 1981). Mathematical receptor models presently in use can be classified into several categories: chemical mass balance (CMB), multivariate, microscopic and source-receptor hybrids. This paper discusses CMB and multivariate models, the two primary types applied in exercises for the Mathematical and Empirical Receptor Models Workshop (Quail Roost II) (Stevens and Pace, 1984). Several new models of both types were presented at the workshop and are not covered by previous reports. To put all these methods of source apportionment into perspective, each model is discussed below. First, receptor models are distinguished from conventional dispersion modeling. Next, various approaches to solving the CMB equation for source contributions are given. Multivariate methods of source apportionment are taken up next. The paper concludes with a discussion of the multicollinearity problem and methods to deal with it.

Dispersion models are the conventional means of predicting the environmental impact of an emission source on air quality. In general, the dispersion model states that the contribution of source j to a receptor, S_j, is the product of an emissions rate, E_j, and a dispersion factor, D_j, so that

$$S_j = D_j E_j. \qquad (1)$$

Dispersion formulae are usually classified by the geometric form of the emission source, giving rise to expressions for point, line, and area sources. A basic goal of air quality meteorology is to relate the dispersion factor for a given source geometry to known meteorological parameters such as wind speed and direction and atmospheric stability (Pasquill, 1974; Benarie, 1980; Seinfeld, 1975). The available dispersion models have been classified and described (U.S.EPA, 1978a, b). Current development of dispersion models focuses on modeling dispersion in rough terrain, long-range transport, wet and dry acidic deposition, and complex atmospheric chemistry.

---

* A Report from the Mathematical and Empirical Receptor Models Workshop (Quail Roost II).

† To whom requests for reprints of this article should be addressed. Requests for reprints of the entire series of six articles (including this one) documenting the Quail Roost II workshop should be directed to Robert K. Stevens, MD-47, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, U.S.A.

In contrast to dispersion models, receptor models start with observed ambient airborne particle concentrations at a receptor and seek to apportion the observed concentrations between several source types based on knowledge of the compositions of the source and receptor materials. Receptor models and their application have been reviewed by Cooper and Watson (1980), Kleinman *et al.* (1980) and Gordon (1980). Core has presented an overview of practical considerations for receptor models in general (1981a) as well as a detailed discussion of CMB methods (1981b).

The foundation of all receptor-based models for particle source assignment is a simple mass conservation argument. If a number of sources, p, exists, and if there is no interaction between their aerosols that causes mass removal or formation, the total airborne particulate mass measured at the receptor, C, will be a linear sum of the contributions of the individual sources S_j

$$C = \sum_{j=1}^{p} S_j. \tag{2}$$

Similarly, the mass concentration of aerosol property i, C_i, will be

$$C_i = \sum_{j=1}^{p} a_{ij} S_j, \tag{3}$$

where a_{ij} is the mass fraction of source contribution j possessing property i *at the receptor*.

CHEMICAL MASS BALANCE MODELS

When the property i is a chemical property, Equation (3) represents a chemical mass balance. If one measures n chemical properties of both source and receptor, n equations of the form of Equation (3) exist. If the number of source types contributing those properties is less than or equal to the number of equations, that is, if $p \leq n$, then the source contributions S_j can be calculated by solving the overdetermined system of linear Equations (3). Five methods of performing this calculation have been applied: the tracer property, linear programming, ordinary linear least-squares fitting, effective variance least-squares fitting, and ridge regression. Each of these methods is discussed below.

Tracer element method

The tracer element method is the simplest. It assumes that each aerosol source type possesses a unique property that is common to no other source type. Equation (3) then reduces to

$$S_j = \frac{C_i}{a_{ij}} \tag{4}$$

for each source j with its own tracer i (using the same notation as in Equation [3]). It works well when the tracers meet the following requirements: (1) a_{ij} perceived at the receptor is well known and invariant between source and receptor, (2) C_i can be measured accurately and precisely in the ambient sample and (3) the concentration of property i at the receptor comes *only* from source type j.

These conditions cannot be completely met in practice, and limiting the model to only one tracer property per source type means that valuable information contained in the other aerosol properties is discarded. However, in some instances, one can use a tracer approach as a first-approximation, upper-limit estimate of source impact. Solutions to the set of Equations (3) have been developed to make use of the additional information provided by more than one chemical property of a source type, and even by other properties (such as particle shape and size). The other methods of solution must deal with the case of a number of constraints, n, greater than the number of unknowns, p.

Linear programming method

Henry (1977b) has applied a linear programming algorithm (Hadley, 1962) that maximizes the sum of the source contributions subject to the following constraints

$$0 \leq S_j \leq M, \text{ and} \tag{5}$$

$$C_i + 3\sigma_{C_i} \geq \sum_{j=1}^{p} a_{ij} S_j, \tag{6}$$

where M is the total particulate mass concentration and σ_{C_i} is the uncertainty in the measurement C_i. However, this solution has not been developed further.

Ordinary linear least-squares method

In the ordinary linear least-squares method (Bevington, 1969), the most probable values of S_j when $n > p$ are achieved by minimizing χ^2, the sum of squares of the differences between the measured values of C_i and those calculated from Equation (3) weighted by σ_{C_i}, the analytical uncertainty of the C_i measurement

$$\chi^2 = \sum_{i=1}^{n} \frac{\left(C_i - \sum_{j=1}^{p} a_{ij} S_j\right)^2}{\sigma_{C_i}^2}. \tag{7}$$

Written in matrix terms, the CMB equation (Equation [3]) is

$$C = AS, \tag{8}$$

where C is the $n \times 1$ vector of observed concentrations, A is the $n \times p$ source composition matrix, and S is the $p \times 1$ vector of source contributions. The ordinary weighted least-squares solution to Equation (8) is

$$S = (A^t W A)^{-1} A^t W C, \tag{9}$$

where W is a diagonal matrix with $\sigma_{C_i}^{-2}$ on the diagonal. Superscript "t" denotes matrix transpose and "-1" is the matrix inverse. This solution provides the added benefit of being able to propagate the measurement uncertainty of C_i through the calculations in order to develop a confidence interval around the calculated S_j. The error estimates of source contributions calculated from ordinary weighted least-squares, assuming errors only in the observed concentrations and no errors in the source composition matrix, are given by Draper and Smith (1981) as the diagonal elements of the matrix

$$(A^tWA)^{-1}. \tag{10}$$

Effective variance least-squares method

The ordinary weighted least-squares solution is incomplete because the ambient aerosol chemical properties C_i are not the only measured observables. The source aerosol chemical properties a_{ij} are also measured observables, but the errors associated with those measurements are not included in the ordinary weighted least-squares fit.

Dunker (1979) and Watson (1979) have applied the treatment of Britt and Luecke (1973) to the solution of a set of Equations (3). The maximum likelihood solution minimizes the function

$$\chi^2 = \sum_{i=1}^{n} \frac{\left(C_i - \sum_{j=1}^{p} a_{ij}S_j\right)^2}{\sigma_{C_i}^2 + \sum_{j=1}^{p} \sigma_{a_{ij}}^2 S_j^2}, \tag{11}$$

where $\sigma_{a_{ij}}$ is the uncertainty associated with the a_{ij} measurement.

The least-squares solution using uncertainties in the A matrix is identical in form to the ordinary weighted least-squares solution in Equation (9) except that the weights $\sigma_{C_i}^{\;2}$ are replaced by

$$V_{\text{eff},i}^{-1} = \left(\sigma_{C_i}^2 + \sum_{j=1}^{p} \sigma_{a_{ij}}^2 S_j^2\right)^{-1}, \tag{12}$$

where V_{eff} is the effective variance.

Since the effective variance solution depends on the source contributions, which are unknown, an iterative procedure is followed. At each iteration, the previously estimated S_j are used to compute new effective variance weights which are used, in turn, to compute new S_j. The process terminates when the S_j do not change much from step to step. The error estimates for the S_j are obtained from the diagonal elements of the matrix

$$(A^tV_{\text{eff}}^{-1}A)^{-1}, \tag{13}$$

where V_{eff}^{-1} is a diagonal matrix with terms from Equation (12) on the diagonal.

This solution provides two benefits. First, it propagates a confidence interval around the calculated S_j which reflects the cumulative uncertainty of all the input observables. The more precise the ambient and source property concentration measurements are, the better the source contribution estimate will be. The second benefit provided by this effective variance weighting is to give those chemical properties with larger measurement uncertainties, or chemical properties that are not as unique to a source type, less weight in the fitting procedure than those properties that are more precisely measured or that are truly unique to a source.

Ridge regression

The CMB Equation (8) can be discussed in statistical terminology as a linear regression problem. In that framework, the observed concentrations C are the dependent variables, the columns of the source composition matrix A are the values of the independent variable, and the source contributions are the regression coefficients to be estimated by least-squares techniques. The following discussion of ridge regression uses this statistical terminology.

An important difficulty is encountered when estimating the separate particulate contributions of sources whose aerosols have similar chemical compositions. This problem, "multicollinearity" is a potentially serious problem for CMB receptor models. Often two sources have nearly identical compositions, or one source composition is almost identical to a linear combination of other source compositions. In this case, the least-squares solution is mathematically unstable, i.e. small errors in the measurements will be magnified into large errors in calculated source contributions. Box *et al.* (1978) present a relatively simple example of the problems caused by instability. Multicollinearity is discussed in more detail at the end of this paper.

Ridge regression is one approach for handling the multicollinearity problem. Another approach is singular value decomposition or regression on principal components (Lawson and Hanson, 1974; Henry and Hidy, 1979). The properties of ridge regression essential to this discussion are briefly summarized below. Ridge regression is discussed further in the statistical literature (Draper and Van Nostrand, 1979; Vinod, 1978; Hoerl and Kennard, 1970; Marquardt, 1970) and in the numerical analysis literature (Lawson and Hanson, 1974).

In ridge regression, a parameter k, the Levenberg–Marquardt smoothing parameter, is introduced into the ordinary least-squares solution

$$S(k) = (A^tWA + kI)^{-1}A^tWC, \tag{14}$$

where $S(k)$ is the ridge solution corresponding to the parameter k, and I is the identity matrix. The effect of k is to assure that the determinant of A^tA is not very small and, therefore, that Equation (14) will be stable. Note that if $k = 0$, the ridge solution becomes the ordinary weighted regression solution given by Equation (9).

The usual statistical practice is to standardize A^tWA to contain correlation coefficients. The S values obtained from Equation (14) are then called standardized coefficients, with the solution in terms of the original data units obtained by an inverse transformation. The diagonal elements of A^tWA in the standardized form are all equal to 1, so a k value of, for example, 0.1

represents a 10% perturbation of the diagonal elements. If $k \neq 0$, a bias or systematic error is introduced into the regression coefficients; the conventional regression solution is unbiased. The expectation is that a large reduction in random error can be achieved at the cost of introducing a small bias in the result. Error estimates for ridge regression are somewhat complicated, but standard results can be found in a paper by Hoerl and Kennard (1970).

Several different automatic approaches for selecting k have been discussed in the literature (Dempster *et al.*, 1977; Wichern and Churchill, 1978). However, none are completely satisfactory, since an automatic criterion for selecting k may perform well in some cases but not in others. Thus, criteria based on physical and statistical judgment must be used after careful consideration. Hoerl and Kennard (1970) describe a statistical approach for determining k. In this approach, k is selected in a stable region in which moderate changes in k do not change the solution significantly.

Weighted regression, effective variance calculations, and ridge regression enhance the ordinary multiple regression approach to the CMB problem. Fortunately, these three features can be used simultaneously or in various combinations. In weighted regression, a transformation is first made to stabilize the error variance of the dependent variable (Draper and Smith, 1981). This transformation can be made using the original error variances of the dependent variable or the effective variances. Subsequently, the regression analysis is performed using the transformed variables; it can be accomplished using either ordinary regression or ridge regression. Finally, the necessary inverse transformation is performed to obtain the solution in the original data units.

In conclusion, it should be noted that methods to reduce instability in CMB calculations such as ridge regression and singular value decomposition have not been extensively tested for this application, and at present should be used with caution.

MULTIVARIATE MODELS

The CMB uses the aerosol property of chemical composition. If source and ambient samples are taken in more than one size range, the size property can also be used to separate the contributions of one source from another. These properties and the CMB results are determined for a single sample. Therefore, the CMB solutions discussed above have no way to incorporate the time variation of ambient concentrations and source emissions. Most of the multivariate methods can do this. Multivariate methods discussed in this section include factor analysis, target transformation factor analysis, extended Q-mode factor analysis, and multiple linear regression, which may be used in conjunction with factor analysis.

While the CMB receptor models are easily derivable from the source model and the elements of their solutions fairly easy to present, multivariate receptor models are not as simple. Watson (1979) has carried through the calculations of the source–receptor model relationship for the correlation and principal component models.

The apparent mathematical complexity of these models does not remove the requirement that every receptor model be representative of and derivable from physical reality as represented by the source model. A statistical relationship between the variability of one observable and another is insufficient to define cause and effect unless this physical significance can be established. The same restriction is imposed on the CMB models.

The multivariate models (except the extended Q-mode model) deal with a series of m measurements of aerosol component i during sampling period or at sampling site k. From Equation (3),

$$C_{ik} = \sum_{j=1}^{p} a_{ij} S_{jk}, \; k = 1 \ldots m. \qquad (15)$$

The multivariate models use C_{ik} with the objective of predicting the number of sources, p, and of predicting which a_{ij} is associated with which S_j (or, more ambitiously, estimating a_{ij} and S_{jk}).

Factor analysis

Factor analysis generally begins with an eigenvector analysis of a cross-product matrix of the data; frequently, the matrix of correlation coefficients

$$r_{C_iC_j} = \frac{1}{m-1} \sum_{k=1}^{m} \left(\frac{C_{ik}-\overline{C}_i}{\sigma_i}\right)\left(\frac{C_{jk}-\overline{C}_j}{\sigma_j}\right), \qquad (16)$$

where m is the number of observations, $\overline{C}_i$ and $\overline{C}_j$ are the average values of C_i and C_j, and σ_i and σ_j are the standard deviations of C_i and C_j. The correlation coefficient is a measure of the extent that ambient concentrations C_i and C_j vary in the same way.

The classical factor analysis model transforms Equation (3) to

$$C_{ik}^{*} = \sum_{j=1}^{p} \alpha_{ij} S_{jk}^{*} + d_j U_{ij}, \qquad (17)$$

where

$$C_{ik}^{*} = \frac{C_{ik}-\overline{C}_i}{(\overline{C_i^2}-\overline{C}_i^2)^{1/2}}, \qquad (18)$$

$$S_{jk}^{*} = \frac{S_{jk}-\overline{S}_j}{(S_j^2-\overline{S}_j^2)^{1/2}}, \qquad (19)$$

and

$$\alpha_{ij} = a_{ij}\left(\frac{S_j^2-\overline{S}_j^2}{C_i^2-\overline{C}_i^2}\right)^{1/2}, \qquad (20)$$

where α_{ij} (the 'factor loading') is related to the source compositions through Equation (20), and S_{jk}^{*} (the 'factor score') is related to the source contributions through Equation (19). The d_j and U_{ij} are unique factor loadings and scores, respectively, and are often left out of the analysis. The result is a principal component analysis

$$C'_{ik} = \sum_{j=1}^{p} \alpha_{ij} S'_{jk}. \tag{21}$$

The factor model treats the C'_{ik} as components of a vector for each chemical component i in an n-dimensional space. If Equation (3) is true, then only p vectors, where $p \leqslant n$ in a less complex p-dimensional space, are required to produce the vectors of C'_{ik}. This p-space is defined by the eigenvectors of the correlation matrix of the C'_{ik}. These p eigenvectors merely define the space, however, and are not necessarily representative of the sources, the S'_j vectors. They must be linearly combined to form the new source vectors. Typically, this is done by applying any one of several algorithms to rotate the axes; one of the most common algorithms is a procedure known as 'VARIMAX' rotation. The classical factor analysis can be used to screen a data set for potential error identification as well as for source recognition (Henry, 1977b; Roscoe *et al.*, 1982).

Target transformation factor analysis

An alternative factor analysis model, target transformation factor analysis (TTFA), has been applied in the apportionment of aerosol mass to sources. In this application, the objective of TTFA is to extract a maximum of information on the number and nature of sources with no or very limited *a priori* information other than the elemental compositional data. The TTFA approach has been detailed by Malinowski and Howery (1980) and applied by Hopke and co-workers (Hopke *et al.*, 1980; Alpert and Hopke, 1980; Hopke, 1981). A different TTFA approach was independently developed by Henry (1977a,b).

One operational difference between TTFA as applied by Hopke and ordinary factor analysis is the use of the correlation about the origin rather than the mean, and between the samples rather than between the elemental species. TTFA uses the cross-product matrix

$$R = (CV^{-1})^t (CV^{-1}), \tag{22}$$

where C is the data matrix, C^t is its transpose and V is a diagonal matrix with elements of the form

$$V_{ii} = \left(\sum_{j=1}^{n} C_{ij}^2 \right)^{-1/2}. \tag{23}$$

This approach permits the development of the source profiles from the data matrix through the target transformation rotation (Roscoe and Hopke, 1981a).

In order to calculate source contributions, it is necessary to rescale the source vectors determined by TTFA so that they are comparable with actual source composition vectors. The necessary scaling factors are determined by multiple regression of total mass against the (unscaled) factor loadings for each sample period. Finally, ordinary weighted least-squares CMB is used to calculate source contributions from the rescaled TTFA source vectors.

Recently, Roscoe and Hopke (1981b) presented a calculational procedure to determine error limits for the elemental composition and mass contribution of each source. This approach permits development (from the data set, without specific profiles for suspected sources) of an approximate source composition matrix. Interpretation of these hypothetical sources as plausible actual sources does, of course, require prior knowledge. The mass contributions are then calculated by an ordinary weighted least-squares approach (Severin *et al.*, 1983).

Multiple linear regression

Midway in complexity between ordinary CMB and factor analysis is the multivariate technique of multiple linear regression (MLR). Like CMB, MLR is based on a linear least-squares fitting process. The procedure requires a tracer element or property to be determined for each source or source category. It is assumed that the amount of this tracer in the sample is proportional to the source strength at the receptor. Next, a MLR is performed on the observed total mass using the tracers as the independent variables. The MLR model is represented mathematically as

$$M_k = b_0 + \sum_{i=1}^{n} b_i T_{ik} + \varepsilon, \tag{24}$$

where M_k is the observed mass on day k, T_{ik} is the observed amount of tracer i on day k, b_0 is the unaccounted-for background mass, b_i is the regression coefficient for tracer i, and ε is random error.

Therefore, on any given day, the amount of mass at the receptor assigned to source i is $b_i T_{ik}$. The MLR method has been applied with success by Kneip *et al.* (1974), Kleinman *et al.* (1980) and Dattner (1978). Note that this model does not assume a known mass fraction of the tracer in each source but determines this implicitly by the multiple regression. Thus, no assumption is made about changes in relative composition between source and receptor. Of course, a general knowledge of which elements are suitable tracers for various sources is required. If large changes in relative composition do occur during transport, the source could become unrecognizable.

The choice of tracer elements or properties is often based on knowledge of emission sources and their compositions. This knowledge can be supplemented by application of factor analysis with a VARIMAX rotation to determine source properties that are highly correlated with a source and also statistically independent of other tracers (Dattner, 1978; Kleinman *et al.*, 1980). It is important that the tracers be more or less independent of each other; otherwise, the MLR will be unstable because of the multicollinearity problem, as explained below.

Advantages and disadvantages of multivariate models

The basic idea of multivariate methods is to use the information contained in the observed interelement correlations and interactions to produce source profiles or factors that can be interpreted as indicative of major source categories. Consequently, if the investigator has a general idea of the expected source compositions, multivariate methods can provide some source resolution even when accurate source composition data are lacking. This characteristic is the major strength of multivariate techniques. In the Quail Roost II exercises, this strength was demonstrated in the ability of several multivariate techniques to uncover the 'unknown' source active in simulated Data Set I (Currie *et al.*, 1984). Therefore, multivariate models are appropriate to use when source profiles are not available or are likely to be very unreliable. They also can be effectively employed in the quality assurance program that examines the data before further analysis (Henry, 1977a,b; Roscoe *et al.*, 1982). These methods can provide a very valuable check on linear least-squares CMB receptor models by alerting the user to the possible presence of unknown sources or possible errors in the data.

To obtain accurate results by multivariate analysis, it is essential that the data set contain an adequate number of samples. An estimate of the minimum number of samples required can be made by determining the number of degrees of freedom per variable. If V variables are measured in N samples, the total number of degrees of freedom in the data set is NV. Each variable's average, standard deviation, and correlation with every other variable are used, in one way or another, in almost any multivariate analysis. These represent relationships among the observed data, each of which decreases by 1 the number of degrees of freedom. Therefore, the total degrees of freedom in a multivariate analysis (for $N > 2$ and $V > 2$) is

$$D = NV - V - \frac{V(V+1)}{2}, \qquad (25)$$

where V is the number of averages and $V(V+1)/2$ is the number of standard deviations and correlations. It is convenient to normalize D by dividing by the number of variables

$$\frac{D}{V} = N - \frac{V+3}{2}. \qquad (26)$$

The experience of one of the authors (RCH) has shown that this number should be at least 30 and preferably 50 to 100. For the requirement $D/V > 30$, the minimum number of samples needed is

$$N > 30 + \frac{V+3}{2}. \qquad (27)$$

In receptor models, the number of species measured is typically 20. For $V = 20$, the minimum acceptable sample size is $N = 42$.

Multivariate methods depend on the variability in the data to deduce the number of major factors or sources in the data. However, the major cause of variability in ambient data is often meteorology, not variations in source emissions. This fact results in highly correlated observations due to the common effect of meteorology on all source contributions at the receptor. In practice, 5–8 factors are generally identified from the data, while there may well be 10 or more important source classes. Thus, factor analysis and similar multivariate techniques are unable to produce a fine resolution of sources. However, they are generally able to find major source classes, single sources with unique composition signatures, or isolated, unusual events that may be significant depending upon the application.

Extended Q-mode factor analysis

This subsection on multivariate methods concludes with a brief description of the use of extended Q-mode factor analysis for source apportionment (Johnson *et al.*, 1984). This methodology differs from the other multivariate models discussed above in that it is applied to a single sample and does not depend on time correlations of concentrations. Thus, Q-mode analysis is really a CMB-type model that uses multivariate methods to deconvolve the receptor composition into a sum of source compositions.

Factor analysis of a data matrix can involve correlating the rows or the columns of the data matrix. The former method is called 'R-mode'; the latter, 'Q-mode.' Thus, 'Q-mode' is a case of pedantic nomenclature (Comrey, 1973). Over a period of years, geologists have developed a number of multivariate techniques based on factor analysis of elemental composition of geological samples. (Some references in the geological literature are: Miesch, 1976a,b; Full *et al.*, 1981). In spirit and application, Q-mode analysis is similar to latent root regression (Webster *et al.*, 1974).

Under an extended Q-mode analysis, the source and receptor composition data are combined into one data matrix. A type of 'correlation' (or cross product about the origin) is calculated from

$$\frac{\sum_k a_{ki} a_{kj}}{\left(\sum_k a_{ki}^2 \sum_k a_{kj}^2\right)^{1/2}}, \qquad (28)$$

where a_{ki} is defined (as before) as the mass fraction of element k in source i.

This cross-product matrix is then subjected to a principal component factor analysis with a VARIMAX rotation. Several specialized computer algorithms (Full *et al.*, 1981) are then applied to determine a set of basic source composition vectors. All other source composition vectors and the receptor composition vector can be expressed as a nonnegative linear sum of these basic source composition vectors. The coefficients of the receptor composition vector expressed as a sum of the basic source vectors represent the contribution (as a fraction of total mass) of each basic source to the receptor mass concentrations. This solution to the CMB equation needs further study and application.

MULTICOLLINEARITY IN SOURCE APPORTIONMENT

Emphasized throughout the reports from Quail Roost II is an important aspect of source apportionment: the handling of sources with similar signatures, that is, sources whose aerosols are chemically and physically similar. Large errors can result if source apportionment by conventional regression or weighted regression analysis is attempted when two or more sources with very similar signatures are included. It is not uncommon for negative aerosol contributions with large magnitudes to be estimated for some sources under these conditions. Another symptom is estimated uncertainties that are larger than the calculated source contributions themselves.

In statistical terms, the problem of similar source signatures is one of 'multicollinearity'. More generally, multicollinearity exists when one source signature is nearly a linear combination of any subset of the other signatures.

Multicollinearities can be identified on the basis of chemical and physical judgment, manual examination of the source signatures, and statistical measures. The first two approaches adequately determine many multicollinearities (for example, sources within the crustal category that have similar signatures). Identifying subtle multicollinearities involving several sources, however, can be more difficult. Thus, the use of statistical measures is also helpful. The variance inflation factor (VIF) is a statistical measure of multicollinearity (Belsley *et al.*, 1980). The VIF quantifies the isolated effect of multicollinearity in terms of the final errors in the source apportionment. Specifically, the VIF can be interpreted as the increase in the error variance of the estimated aerosol contribution of a specific source due to multicollinearity alone. Also, principal component analysis can be used to explicitly uncover multicollinearities in the source composition matrix (Gunst and Mason, 1977; Henry, 1982).

One approach to the multicollinearity problem is to examine the source signatures and determine groups of sources whose aerosols are chemically similar. One such group, for example, could include sources in the crustal category, such as soil and road dust. Then a signature could be selected that represents the entire group. Subsequently, an analysis could be made to apportion the aerosols among the selected groups of sources. This approach (source selection) reduces the multicollinearity problem and the potentially large errors that often are caused by it. However, grouping sources reduces the resolution of the source apportionment.

Source selection may offer a solution or a partial solution, depending on the situation. It may be possible to eliminate some sources on the basis of engineering judgment. A source should be eliminated on this basis only if there is strong physical evidence (such as wind direction data for the sampling period in question) that it does not contribute significantly to the monitored aerosol concentrations. Deleting sources may eliminate some of the multicollinearity problems. Source selection can be made on the basis of emissions inventories, source–receptor geometry, wind information, additional aerosol characterization, etc.

One fundamental problem with eliminating multicollinearity by source selection involves determining realistic error estimates for those sources that remain in the least-squares solution. Their uncertainties should reflect the possibility that the wrong sources were selected; currently, this is not the case. A further difficulty arises in assigning uncertainties to source contribution averages. If a source is not selected on one day, it is usually averaged in as 0, with an uncertainty of 0. The error of the average can then be determined by propagation of the errors on individual source contributions. In this case, the uncertainty on the average is given by $N^{-1}(\Sigma\sigma_i^2)^{1/2}$, where N is the number of sampling periods and σ_i^2 is the error variance associated with the CMB source contribution determination for each period. Another approach that has been applied is to take the uncertainty of the average source contribution to be $\sigma N^{-1/2}$, where N is as before but σ is the standard deviation of the N source contributions. These two approaches are not equivalent, and further work is needed to better treat the question of uncertainties of average source contributions.

Another approach to the problem is the use of statistical techniques specifically designed to handle multicollinearity (e.g. ridge regression and singular value decomposition). Source apportionment through ridge regression was discussed earlier in this report. An additional approach is regression on principal components (Henry and Hidy, 1979). Depending on the situation, a combination of the approaches discussed here may be needed. Without additional information, such as wind direction, no CMB or multivariate receptor model can distinguish between two sources whose signatures are indistinguishable.

Finally, a fundamental analysis is needed to clarify the calculation of uncertainties in receptor models and to develop clear guidance for unsophisticated users on how to handle multicollinearity problems in practical situations. Henry (1982) has taken a preliminary step in this direction. In that paper, a fundamental limit is determined for the number of sources that can be determined to a given level of accuracy by a weighted least-squares CMB calculation for a given source composition matrix. Henry (1982) also presents a method for optimal CMB estimation of linear combinations of sources that cannot be accurately estimated separately.

REFERENCES

Alpert D. J. and Hopke P. K. (1980) A quantitative determination of the sources in the Boston urban aerosol. *Atmospheric Environment* 14, 1137 1146.

Belsley D. A., Edwin K. and Welsch R. E. (1980) *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.* John Wiley, New York.

Benarie M. M. (1980) *Urban Air Pollution Modeling.* MIT Press, Cambridge, Massachusetts.

Bevington P. R. (1969) *Data Reduction and Error Analysis for the Physical Sciences.* McGraw-Hill, New York.

Box G. E. P., Hunter W. G. and Hunter J. S. (1978) *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building,* pp. 489–493. John Wiley, New York.

Britt H. I. and Luecke R. H. (1973) The estimation of parameters in non-linear, implicit models. *Technometrics* **15,** 233–247.

Comrey A. L. (1973) *A First Course in Factor Analysis.* Academic Press, New York.

Cooper J. A. and Watson J. G. (1980) Receptor oriented methods of air particulate source apportionment. *J. Air Pollut. Control Ass.* **30,** 1116–1125.

Core J. E. (1981a) *Receptor Model Technical Series, Vol. I: Overview of Receptor Model Applications to Particulate Source Apportionment.* EPA-450/4-81-016a, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Core J. E. (1981b) *Receptor Model Technical Series, Vol. II: Chemical Mass Balance.* EPA-450/4-81-016b, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Currie L. A., Gerlach R. W., Lewis C. W., Balfour W. D., Cooper J. A., Dattner S. L., De Cesar R. T., Gordon G. E., Heisler S. L., Hopke P. K., Shah J. J., Thurston G. D. and Williamson H. J. (1984) Interlaboratory comparison of source apportionment procedures: results for simulated data sets. *Atmospheric Environment* **18,** 1517–1537.

Dattner S. L. (1978) *Preliminary Analysis of the Use of Factor Analysis of X-ray Fluorescence Data to Determine the Sources of Total Suspended Particulate.* Feasibility Study, Texas Air Control Board, Austin, Texas.

Dempster A. P., Schatzoff M. and Wermuth N. (1977) A simulation study of alternatives to ordinary least-squares. *J. Am. Statist. Ass.* **72,** 77–106.

Draper N. R. and Smith H. (1981) *Applied Regression Analysis.* John Wiley, New York.

Draper N. R. and Van Nostrand R. C. (1979) Ridge regression and James–Stein estimation: review and comments. *Technometrics* **21,** 451–465.

Dunker A. M. (1979) *A Method for Analyzing Data on the Elemental Composition of Aerosols.* Research Publication GMR-3074ENV-67, General Motors Research Laboratory, Warren, Michigan.

Full W. E., Ehrlich R. and Klovan J. E. (1981) Extended Q model—objective definition of external end members in the analysis of mixtures. *Math. Geol.* **13,** 331–344.

Gordon G. E. (1980) Receptor models. *Envir. Sci. Technol.* **14,** 792–800.

Gunst R. F. and Mason R. L. (1977) Advantages of examining multicollinearities in regression analysis. *Biometrics* **33,** 249–260.

Hadley G. (1962) *Linear Programming.* Addison Wesley, Reading, Massachusetts.

Henry R. C. (1977a) The application of factor analysis to urban aerosol source identification. In *Proc. Fifth Conf. on Probability and Statistics in Atmospheric Sciences,* American Meteorological Society, Las Vegas, Nevada, pp. 134–138.

Henry R. C. (1977b) A factor model of urban aerosol pollution. Ph.D. dissertation, Oregon Graduate Center, Beaverton, Oregon.

Henry R. C. (1982) Stability analysis of receptor models that use least-squares fitting. Presented at the Specialty Conference on Receptor Models (Air Pollution Control Association), 18–20 October, Danvers, Massachusetts.

Henry R. C. and Hidy G. M. (1979) Multivariate analysis of particulate sulfate and other air quality variables by principal components—I. Annual data from Los Angeles and New York. *Atmospheric Environment* **13,** 1581–1596.

Hoerl A. E. and Kennard R. W. (1970) Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* **12,** 55–67.

Hopke P. K. (1981) The application of factor analysis to urban aerosol source resolution. In *Atmospheric Aerosol: Source/Air Quality Relationships* (edited by Macias and Hopke), pp. 21–49. ACS Symp. Series No. 167, American Chemical Society, Washington, D.C.

Hopke P. K., Lamb R. E. and Natusch D. F. S. (1980) Multi-elemental characterization of urban roadway dust. *Envir. Sci. Technol.* **14,** 164–172.

Johnson D. L., Davis B. L., Dzubay T. G., Hasan H., Crutcher E. R., Courtney W. J., Jaklevic J. M., Thompson A. C. and Hopke P. K. (1984) Chemical and physical analyses of Houston aerosol for interlaboratory comparison of source apportionment procedures. *Atmospheric Environment* **18,** 1539–1553.

Kleinman M. T., Pasternack S., Eisenbud M. and Kneip T. J. (1980) Identifying and estimating the relative importance of sources of airborne particulates. *Envir. Sci. Technol.* **14,** 62–65.

Kneip T. J., Kleinman M. T. and Eisenbud M. (1974) Relative contribution of emission sources to the total airborne particulates in New York City. In *Proc. Third Int. Clean Air Congress,* Düsseldorf, West Germany, pp. C96–C97.

Lawson C. L. and Hanson R. J. (1974) *Solving Least Squares Problems.* Prentice-Hall, Englewood Cliffs, New Jersey.

Malinowski E. R. and Howery D. G. (1980) *Factor Analysis in Chemistry.* John Wiley, New York.

Marquardt D. W. (1970) Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics* **12,** 591–612.

Miesch A. T. (1976a) Q-mode factor analysis of compositional data. *Comput. Geosci.* **1,** 147–159.

Miesch A. T. (1976b) Interactive computer programs for petrologic modeling with extended Q-mode factor analysis. *Comput. Geosci.* **2,** 439–492.

Pasquill F. (1974) *Atmospheric Diffusion,* 2nd Ed. Halsted Press, John Wiley, New York.

Roscoe B. A. and Hopke P. K. (1981a) Comparison of weighted and unweighted target transformation rotations in factor analysis. *Comput. Chem.* **5,** 1–7.

Roscoe B. A. and Hopke P. K. (1981b) Error estimates for factor loadings and scores obtained with target transformation factor analysis. *Analytica chimica Acta* **132,** 89–97.

Roscoe B. A., Hopke P. K., Dattner S. L. and Jenks J. M. (1982) The use of principal component factor analysis to interpret particulate compositional data sets. *J. Air Pollut. Control Ass.* **32,** 637–642.

Seinfeld J. H. (1975) *Air Pollution: Physical and Chemical Fundamentals.* McGraw-Hill, New York.

Severin K. G., Roscoe B. A. and Hopke P. K. (1983) The use of factor analysis in source determination of particulate emissions. *Particulate Sci. Technol.* **1,** 183–192.

Stevens R. K. and Pace T. G. (1984) Overview of the Mathematical and Empirical Receptor Models Workshop (Quail Roost II). *Atmospheric Environment* **18,** 1499–1506.

U.S. Environmental Protection Agency (1978a) *Workbook for Comparison of Air Quality Models.* OAQPS Guideline Series. EPA-450/2-78-028a, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

U.S. Environmental Protection Agency (1978b) *Workbook for Comparison of Air Quality Models – Appendices.* OAQPS Guideline Series. EPA-450/2-78-028b, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Vinod H. D. (1978) A survey of ridge regression and related techniques for improvements over ordinary least-squares. *Rev. Econ. Statist.* **60,** 121–131.

Watson J. G. (1979) Chemical element balance receptor model methodology for assessing the sources of fine and total suspended particulate matter in Portland, Oregon

Reproduced from
best available copy.

Ph.D. dissertation, Oregon Graduate Center, Beaverton, Oregon.

Watson J. G. *et al.* (1981) *Receptor Models Relating Ambient Suspended Particulate Matter to Sources*. EPA-600/2-81-039. U.S. Environmental Protection Agency, Researc., Triangle Park, North Carolina.

Webster J. T., Gunst R. F. and Mason R. L. (1974) Latent root regression analysis. *Technometrics* **16**, 513–522.

Wichern D. W. and Churchill G. A. (1978) A comparison of ridge estimators. *Technometrics* **20**, 301–311.

Atmospheric Environment Vol. 18, No. 8, pp. 1517–1537, 1984
Printed in Great Britain

0004-6981/84 $3.00 + 0.00
Pergamon Press Ltd.

INTERLABORATORY COMPARISON OF SOURCE APPORTIONMENT PROCEDURES: RESULTS FOR SIMULATED DATA SETS*†

LLOYD A. CURRIE,‡ ROBERT W. GERLACH
National Bureau of Standards, Washington, DC 20234, U.S.A.

CHARLES W. LEWIS
U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, U.S.A.

W. DAVID BALFOUR
Radian Corporation, Austin, TX 78766, U.S.A.

JOHN A. COOPER
NEA, Inc., Beaverton, OR 97005, U.S.A.

STUART L. DATTNER
Texas Air Control Board, Austin, TX 78723, U.S.A.

RICHARD T. DE CESAR
Oregon Graduate Center, Beaverton, OR 97005, U.S.A.

GLEN E. GORDON
Department of Chemistry, University of Maryland, College Park, MD 20742, U.S.A.

STEVE L. HEISLER
Environmental Research and Technology, Inc., Concord, MA 01742, U.S.A.

PHILIP K. HOPKE
Institute for Environmental Studies, University of Illinois at Urbana-Champaign, Urbana, IL 61801, U.S.A.

JITENDRA J. SHAH
Environmental Research and Technology, Inc., Concord, MA 01742, U.S.A.

GEORGE D. THURSTON
School of Public Health, Harvard University, Boston, MA 02115, U.S.A.

and

HUGH J. WILLIAMSON
Radian Corporation, Austin, TX 78766, U.S.A.

(*First received 9 June* 1983 *and in final form* 19 *January* 1984)

Abstract—Three sets of simulated aerosol compositional data were prepared to (1) assess the current state of the art of source apportionment procedures and (2) provide initial sets of test data to aid method development. The data sets were generated from reported source profiles, real meteorological data (St. Louis, 1976), and two constructed city plans. Following plume dispersion by means of the RAM model, 40 'samples' having known source contributions and error structure were generated for each set. Seven laboratories participating in the Mathematical and Empirical Receptor Models Workshop (Quail Roost II) undertook

---

[*] A publication of the National Bureau of Standards, not subject to copyright.

[†] A Report from the Mathematical and Empirical Receptor Models Workshop (Quail Roost II)

[‡] To whom requests for reprints of this article should be addressed. Requests for reprints of the entire series of six articles (including this one) documenting the Quail Roost II workshop should be directed to Robert K. Stevens, MD-47, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, U.S.A.

deconvolution of one or two of the sets by various numerical techniques (chemical mass balance or multivariate). Comparison of the participants' results with the known source contributions showed that the source contribution estimates were consistent with the truth within a factor of ~ 2 and that the uncertainty estimates ranged from much too conservative (broad) to much too small. No unique method of choice emerged from this exercise; the participants' various techniques appeared complementary and capable of resolving ~ 6–9 different sources. The intercomparison did allow us to formulate suggestions for improving the simulation process *per se* and for improving the various treatments (especially with respect to estimating uncertainties).

Key word index: Atmospheric particles, chemical mass balance, error propagation, factor analysis, laboratory intercomparison, multivariate analysis, simulation data, source apportionment, uncertainty estimates.

INTRODUCTION

Interlaboratory comparison represents a crucial step in the validation or improvement of analytical measurement processes, particularly as applied to difficult, 'real-world' problems involving multiple species at the trace level. Almost without exception, such intercomparisons reveal large discrepancies among laboratories—discrepancies that may arise from errors in sampling, chemical measurements, or data evaluation models and methods. Important insight is often gained by separately examining these three steps of the overall measurement process. Assessment of the state of the art of the last step (data evaluation) is the objective of this article.

The evaluation and reporting of analytical data can be separately investigated, in the interlaboratory environment, through the medium of a simulation exercise in which the system under investigation is modeled by computer. This approach permits one to more or less realistically simulate the physical (chemical, meteorological, etc.) processes that occur in the field, while maintaining full control over the system model and its error structure. That is, the 'truth', including the magnitude and distribution of the individual random and systematic errors, is fully known (to those who generate the data). Such was the stimulus and the spirit of the present interlaboratory exercise. It was modeled, in part, from the International Atomic Energy Agency's very well crafted and very informative gamma ray deconvolution interlaboratory comparison using convolved, simulated gamma ray spectra (Parr *et al.*, 1979).

Within the context of source apportionment analysis, a simulated data set is a collection of mathematically generated numbers that are intended to represent a set of ambient air quality measurements (typically, aerosol elemental concentrations). In contrast to a 'real' data set resulting from actual field measurements, however, in a simulated data set the impacts of the contributing sources are known exactly from knowledge of the data generation process. It is this absolute standard of truth against which predictions of various receptor models may be compared that makes the use of simulated data sets such a potentially powerful (and prior to now, little used) method to evaluate the performance of alternative mathematical source apportionment methods. This report describes the generation of the simulated data sets made available to participants of the Mathematical and Empirical Receptor Models Workshop (Quail Roost II) (Stevens and Pace, 1984), and the source apportionment results based on the participants' analysis of those data sets. While many different aspects of the participants' results are presented, Tables 12 and 13 and Fig. 5 give the key results.

The simulated data sets produced for Quail Roost II are expected to be useful well beyond the original application for which they were created. Despite their various limitations (discussed later in this report), they are the product of the most comprehensive attempt to date to realistically model the ambient elemental concentrations that could be plausibly expected from a distribution of commonly occurring sources. The data sets should be useful for testing new or modified receptor models, as well as for debugging existing receptor model computer programs. The data sets are in fact already being used for these purposes at New York University, the University of Maryland, the National Bureau of Standards, the State University of New York College of Environmental Science and Forestry at Syracuse, and the U.S. Environmental Protection Agency.

GENERATION OF SIMULATED DATA SETS

Three data sets were constructed as outlined in Table 1. Each source profile consisted of 19 elements and the isotopic ratio $^{14}C/^{12}C$ (not provided in Data Set I), with component concentrations ($mg\ g^{-1}$) meant to be representative of a fine-fraction aerosol component. Using contrived source geographies (Figs 1 and 2) and the RAM Gaussian plume dispersion model Version 80336 (Novak and Turner, 1976), ambient concentrations of each species were generated for each of forty 12-h averaging periods. All simulations were run with real meteorological data obtained from the National Climatic Center (Asheville, NC). Starting with Julian day 249 of 1976, hourly surface data from St. Louis/Lambert, MO, were combined with twice-daily mixing height information from Salem, IL, in the meteorological preprocessing program RAMMET (Turner and Novak, 1978) to obtain hourly wind speeds, wind directions, stability classes, mixing heights, and air temperatures. These output quantities

Table 1. Structure of simulated data sets

Generating equation

$$C_i^{(K)} = \sum^{p} [A - e_m + e_H^{(K)}]_{ij} S_j^{(K)} + e_i^{(K)}$$

where:

p = number of active sources ($p \leqslant 13$)

K = sampling period ($1 \leqslant K \leqslant 40$)

$C_i^{(K)}$ = "observed" concentration of species i for period K ($1 \leqslant i \leqslant N, N \leqslant 20$)

$S_j^{(K)}$ = true intensity (at receptor) of source $j (1 \leqslant j \leqslant p)$

A_{ij} = 'observed' source profile matrix (element i, j)

e_i = random measurement errors, independent and normally distributed

e_m = systematic source profile errors, independent and normally distributed (systematic because fixed over the 40 sampling periods)

e_H = random source profile variation errors, independent and log-normally distributed

Data set characteristics

Set I: $p = 9$ (including one unknown source);* errors = e_i, e_m; City Plan No. 1 (Fig. 1)

Set II: $p = 13$ (all known); errors = e_i, e_m; City Plan No. 2 (Fig. 2)

Set III: $p = 13$ (all known); errors = e_i, e_m, e_H; City Plan No. 2 (Fig. 2)

* For Data Set I, participants were told only that $p \leqslant 13$.

were then employed in the generation of ambient concentrations by the dispersion model. The original unprocessed meteorological information was provided to the participants. An additional component of source variation in the model, not provided to participants, was the hourly source strength fluctuation due to daily operating schedules created for each classification of sources. For example, type A and B steel plants followed the same diurnal pattern based on a weekly schedule.

Expressed in matrix form, the generating equation from Table 1 is

$$C = AS + e, \tag{1}$$

where the vector of 'observed' concentrations (ng m^{-3}) is represented by C, the matrix of source profiles (mg g^{-1}) by A, and the source contributions (μg m^{-3}) by S. As Table 1 indicates, each data set included normally distributed measurement errors for the individual observations (e_i, random) and for the source

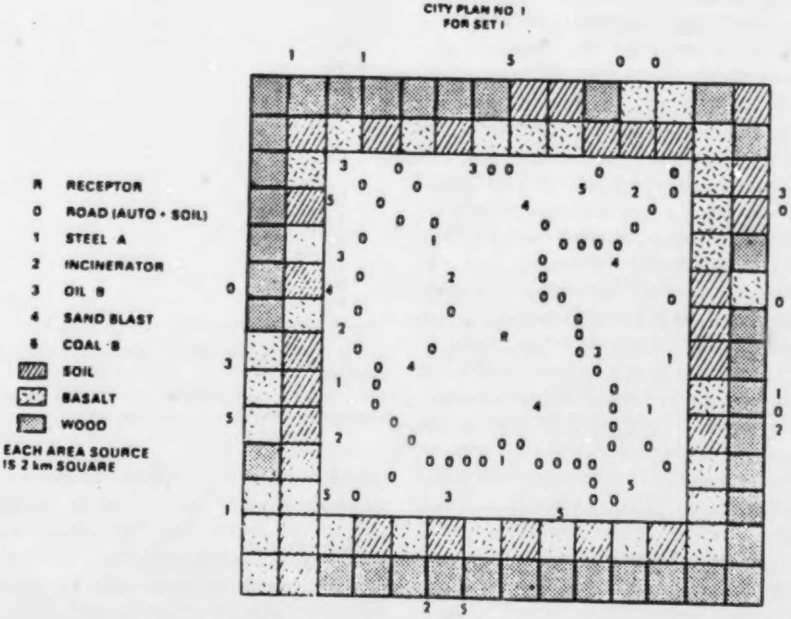

Fig. 1. City Plan No. 1 (for Data Set I).

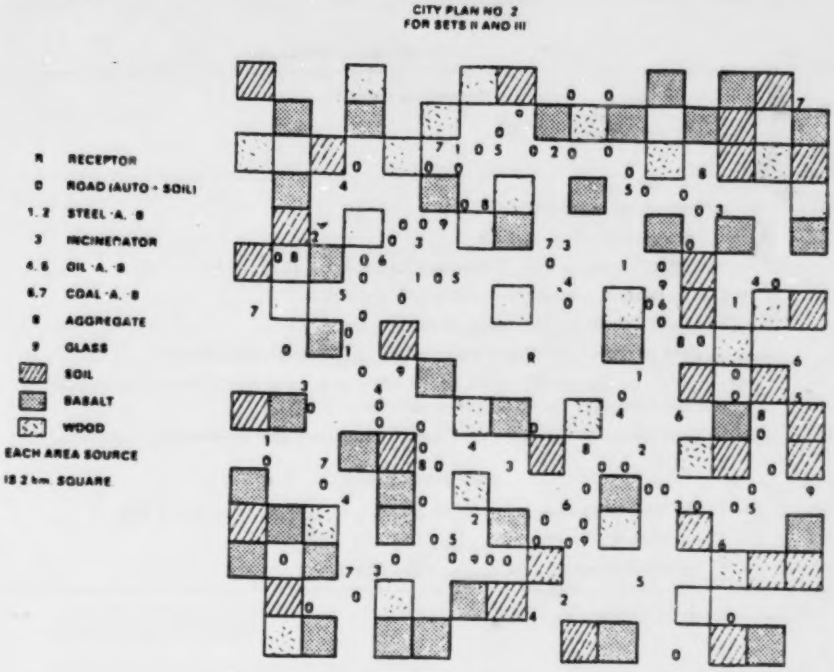

Fig. 2. City Plan No. 2 (for Data Sets II and III).

profiles (e_m, systematic). Set III also included log-normal profile errors (e_H, random) for each of the sources. In other words, the profiles themselves changed from one period to the next. In all cases participants were informed of the distributions and standard deviations of the error components (Tables 2 and 3). A 'world list' of 13 source profiles was also provided. All of these profiles were active in Sets II and III, but only 8 were active in Set I. In addition, Set I included one source type, 'sandblast', of which the participants were not informed. The simulated data sets thus had a well-defined structure in terms of the linear model and error terms. The sets were realistic because mixing of source emissions at the receptor was accomplished via dispersion, and random and systematic errors were added in accordance with typical field values (although the selected e_H errors were somewhat smaller than what realism would probably require). The data and model did assume, however, that all species were conserved from source to receptor.

Some deficiencies of the dispersion model employed are noted later in this report. It is important to emphasize that any such deficiencies have no effect on the *accuracy* of the resulting simulated data sets *per se*. The dispersion model is employed to link source emissions, locations and meteorology in order to assign exact numerical values to the source intensities S_j at the receptor site. These intensities are then convoluted and perturbed with random errors to produce the "observed" ambient elemental concentrations C_i according to Equation (1). The job of any source apportionment analysis is then to [illegible] the S_j values from the C_i values. Dispersion model deficiencies *do* affect the '*realism*' of the simulated data sets to the degree that the deficiencies distort the relationships among the data set values and the emissions inventory, and among the data set values and the meteorological conditions and source 'geography'.

Table 2. 'Measurement' errors* for simulated data sets

Element	Relative error (%)	Absolute error (ng m^{-3})
C	8	200
Na	10	200
Al	15	100
Si	15	50
S	10	20
Cl	10	5
K	10	15
Ca	14	10
Ti	8	7
V	8	2
Cr	8	3
Mn	8	1
Fe	5	3
Ni	10	1
Cu	10	3
Zn	10	2
As	10	12
Br	8	1
Pb	8	1
CC†	8	5%

*Quoted measurement error = absolute error + (elemental concentration × relative error)/100.

†Contemporary carbon (determined from ^{14}C concentration) as a percentage of total carbon.

Table 3. Random errors* in source profiles for Set III

Element	Steel-A	Steel-B	Oil-A	Incin-erator	Oil-B	Coal-A	Aggre-gate	Glass	Basalt	Soil	Wood	Coal-B
C	1.2	1.2	1.2	1.2	1.2	1.2	1.5	1.5	1.2	1.2	1.3	1.2
Na	1.3	1.3	1.2	2.0	1.2	2.0	1.5	1.3	1.5	1.5	2.5	2.0
Al	1.3	1.3	1.1	3.0	1.1	1.2	1.1	2.5	1.1	1.1	3.0	1.2
Si	2.0	2.0	1.3	2.5	1.3	1.3	1.1	1.2	1.1	1.1	3.0	1.3
S	1.5	1.5	1.1	3.0	1.1	2.0	1.2	1.3	1.2	1.2	1.2	2.0
Cl	1.3	1.3	3.1	1.3	3.1	1.3	1.3	1.1	1.3	1.3	2.0	1.3
K	2.0	2.0	1.5	1.5	1.5	2.0	1.5	1.3	1.5	1.5	1.5	2.0
Ca	2.0	2.0	3.0	2.0	3.0	1.2	1.2	1.1	1.2	1.2	1.2	1.2
Ti	2.0	2.0	1.5	3.0	1.5	1.3	1.2	1.2	1.2	1.2	2.0	1.3
V	1.5	1.5	1.5	3.0	1.5	2.0	1.1	1.1	1.2	1.2	2.0	2.0
Cr	2.5	2.5	2.0	3.0	2.0	3.0	1.3	3.0	1.3	1.3	2.0	3.0
Mn	2.0	2.0	1.3	3.0	1.3	2.5	1.3	2.5	1.2	1.2	2.0	2.5
Fe	1.5	1.5	1.2	2.5	1.2	1.5	1.3	3.0	1.3	1.3	2.0	1.5
Ni	1.5	1.5	2.0	2.0	2.0	2.0	1.3	3.0	2.0	2.0	2.5	2.0
Cu	2.0	2.0	2.0	1.5	2.0	2.0	1.3	2.0	1.5	1.5	1.5	2.0
Zn	2.0	2.0	2.0	1.3	2.0	3.0	1.3	3.0	1.2	1.2	1.3	3.0
As	1.5	1.5	3.0	1.5	3.0	1.5	1.2	3.0	1.1	1.1	2.0	1.5
Br	1.5	1.5	3.0	1.5	3.0	1.5	1.5	3.0	1.5	1.5	1.5	1.5
Pb	2.5	2.5	2.5	1.5	2.5	3.0	2.0	2.5	2.0	2.0	1.2	3.0
CC†	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1

* Errors have a log-normal distribution whose standard deviation is given by $\log_{10}$ (tabulated value).
† Contemporary carbon (determined from ^{14}C concentration) as a percentage of total carbon.

These aspects were generally unimportant in the present case because the source apportionment strategies presented at Quail Roost II made little direct use of the emissions inventories and meteorological information that were supplied. 'Realism' also may have suffered somewhat as a result of the large numbers of duplicate sources (especially with City Plan No. 2). This duplication was intended partly to compensate for the narrowness of the dispersion plume, and tended to decrease the variability of source contributions from period to period. (See Tables 5A–C for the actual variations in source contributions.)

The complexity of the data sets increased in the order I → II → III, because of the increased number of active components in the latter two sets and especially because of the considerable log-normal variability in source profiles in Set III. In many respects, however, Set I was the most interesting because participants were not told how many sources from the world list were active, and because an additional totally unknown source was also present. In summary, the simulated data sets were believed to comprise fairly realistic but well controlled sets of test data for reliable evaluation of precision and accuracy in the final data evaluation step (deconvolution) of the source apportionment process.

To illustrate the general nature and level of difficulty of the simulated data, several highlights of Set I are given in Fig. 3. Here, the 13 source types given to the participants are indicated in the first column, together with 2 additional source types not sent to the participants. The first additional source type, 'road', was simply a linear combination of the known source types 'auto' (75%) and 'soil' (25%) and was used to represent the reentrainment of road dust by motor vehicles. While soil was used on its own as one of the sources in the data set generation, auto was used only in the road combination. Thus, the auto source intensity is simply 75% of the road intensity, whereas the soil intensity is the sum of 25% of the road intensity and a quantity independent of the road source. This method of using soil, auto and road was the same in the generation of all three data sets. The second additional source type, 'sandblast', was a source totally unknown to the participants.

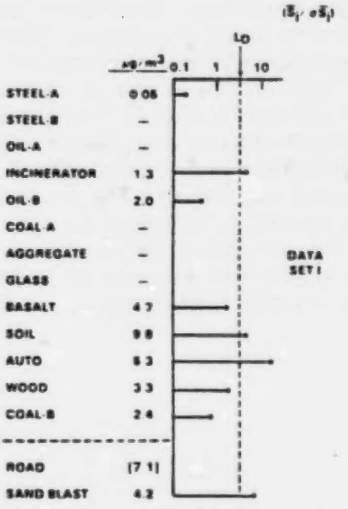

Fig. 3. Characteristics of Data Set I.

Average concentrations of the 9 active sources at the receptor, given in the second column of Fig. 3, were adjusted during the data set generation step so that a range of difficulty (or detectability) would be represented, as indicated by the $\overline{S}_j/\sigma_{\overline{S}_j}$ bar graph. The theoretical (signal/noise) detection limit of 3.29 (Currie, 1977) is indicated by the arrow on the horizontal axis. The weakest source, steel-A, was well below the detection limit. However, several sources (including the unknown sandblast) were easily detectable, since the true average concentrations exceeded by many times the standard errors that would result from weighted least-squares deconvolution of the average elemental concentrations.

The intensity of each source type varied greatly over the 40 averaging periods. Figure 4 depicts the range of intensity variation due to the incinerator source at the receptor site in Set I over all sampling periods. For this typical source, a large fraction of the intensities exceeded the (approximate, weighted least-squares) detection limit (L_D in the figure). The rather sharp variations seen in this plot are due to the unrealistically narrow dispersion characteristics of the RAM model. Minimizing the undesirable effect of this narrowness on the resulting data sets was a primary factor in planning the number of source locations for each source type. Even so, this artifact of the RAM model was never completely overcome. Thus the resulting distributions of each source contribution as a function of concentration were noticeably skewed, as can be inferred from Fig. 4, for example. This effect has been previously characterized by Bowne *et al.* (1981).

Table 4 lists the true source profiles from which the data sets were generated. The participants were provided not with these, however, but with a similar set of profiles perturbed by the systematic errors e_m. (The participants were not given the sandblast or road source profiles.) Tables 5(A) (C) list the true mass contributions by source type and observation period for all three data sets. These contributions were, of course, not given to the participants. Table 6 lists the emissions inventory given to the participants, which was input to the RAM dispersion model for generation of Sets II and III. The primary consideration in choosing emission rates was to produce (in conjunction with the dispersion model) elemental ambient concentrations that were consistent with common experience. That this consideration required unrealistically high emission rates is a direct indication of limitations of the RAM model. Our goal of matching 'common experience', however, was not fully achieved. For example, the contributions of the crustal components (soil, basalt) exceeded what is commonly found in the fine-fraction aerosol in U.S. cities. The three data sets themselves, analyzed by the participants, are not included in this article because of their length.

SOURCE APPORTIONMENT SOLUTION STRATEGIES

Several distinct approaches were chosen by individual participants in their attempts to estimate average source intensities $\overline{S}_j$ as well as uncertainties from one or more of the simulated data sets. Each of these approaches is summarized below. The form of each summary is an abbreviated series of steps followed by a narrative giving details and illustrations of some of the steps. This method of presentation is intended to show the basic structure of each approach as simply as possible, to reveal which portions are 'recipe-like', and which portions require experienced judgment. One measure of uncertainty, σ_j (i.e. $\sigma_{\overline{S}_j}$), the standard error in the estimated source impact $\overline{S}_j$, was reported by

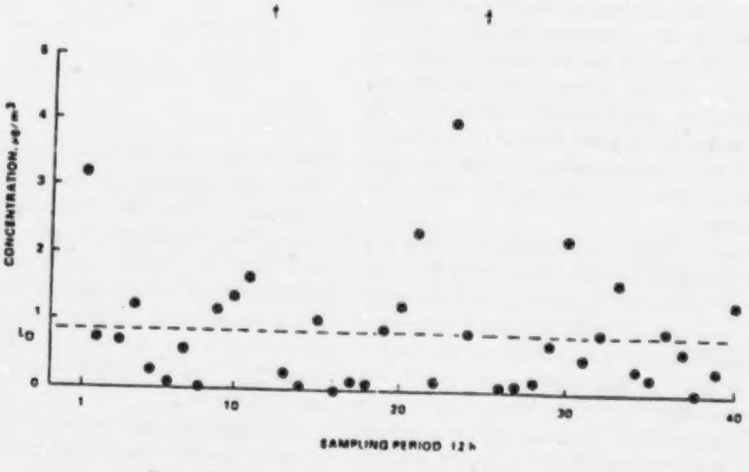

Fig. 4 Incinerator impact at receptor for Data Set I

Table 4. True source profiles (mg g^{-1})

Element	Steel-A	Steel-B	Oil-A	Incinerator	Oil-B	Coal-A	Aggregate	Sandblast	Glass	Basalt	Soil	Auto	Wood	Coal-B	Road
C	150	210	37	40	200	30	1	5	2	0.1	66.4	550	600	8	429
Na	11	14	25	120	22	9.7	6	8	150	19	4	70	3	4	52.5
Al	13	10	7	4.2	5.4	26	70	7.5	3.6	88	222	8.6	7	140	62
Si	8	12	180	7.0	70	150	300	400	2.5	240	543	15	8.9	200	147
S	27	48	140	130	110	15	7.5	8.8	170	0.3	0.3	8	1	180	6.1
Cl	30	34	0.9	270	12	11	0.8	28	2.7	0.2	0.1	10.7	10	17	8.1
K	9.1	13	0.9	200	1.4	10	14	7	27	8.3	12	2.7	8	5	5
Ca	70	45	0.9	11	8.6	23	40	8.2	4.6	67	20.5	11.8	9	40	13.9
Ti	1.3	1.0	0.7	0.9	1.1	2.6	2.8	4	0.4	9	12.7	0	0.02	9	3.2
V	1.8	1.3	9.2	0.03	0.7	0.1	0.07	0.03	1.5	0.2	0.3	0.1	0	0.8	0.1
Cr	3.3	3.2	0.2	0.9	0.4	0.4	0.06	5.8	3.4	0.1	0.3	0.09	0	0.3	0.1
Mn	16	22	0.1	0.3	0.1	0.3	0.8	1.4	0.3	1.5	2	0.6	0.02	0.2	0.9
Fe	120	160	14	6.1	13	18	53	63	10	86	97	7	1	70	29.5
Ni	2.8	2.9	4.9	0.1	0.06	0.2	0.03	3.2	1.3	0.1	0.1	0.06	0.01	0.6	0.07
Cu	2.6	3.7	0.7	3.6	0.8	1.3	0.6	6.9	0.4	0.1	0.2	0	0	0.4	0.05
Zn	25	52	0.4	26	1.3	3	0.4	4.7	0.7	0.08	0.4	3.2	0.7	0.9	2.5
As	0.08	0.1	0.02	0.2	0.3	0.3	0.03	0.1	0.6	0	0.01	0.01	0	0.2	0.01
Br	0.1	0.1	0.03	0.8	0.4	0.2	0.01	0.05	0.02	0	0.01	41.7	0.08	0.01	31.3
Pb	7.6	14	0.8	17	3.8	2	0.1	5.8	0.5	0.01	0.3	120	0.1	1.2	90.1
CC*	0	0	0	50	0	0	0	0	0	0	97	2	108	0	6

* Contemporary carbon (determined from ^{14}C concentration) as a percentage of total carbon. Values may exceed 100% because of the age of the wood and atmospheric nuclear testing.

Table 5A. True mass contributions (ng m^{-3}) by source type and observation period for Set 1

Period	Steel-A	Incinerator	Oil-B	Sandblast	Basalt	Soil	Wood	Coal-B	Road
1	39	3200	48	7827	2767	5153	2272	498	9464
2	260	747	7253	652	2374	6400	7032	674	3579
3	96	717	10	976	3726	1999	3486	924	3033
4	168	1244	9780	855	5298	2014	8082	43	4292
5	12	250	3	1861	4554	11558	3333	5233	4013
6	2	61	596	2139	6586	8179	8896	1244	3644
7	25	578	64	3462	13921	6589	1899	2387	4704
8	50	18	11	14565	5646	6859	3607	1295	18
9	84	1154	608	1	1613	805	484	298	16
10	147	1384	888	0	722	2475	6218	3332	2768
11	11	1663	2416	3024	4360	4232	1552	892	21861
12	122	8622	4039	200	12216	8779	8393	15873	138
13	28	279	78	3672	8650	19077	425	2406	1521
14	0	49	358	17246	4456	8183	3057	1678	935
15	22	1048	5	3683	3608	8968	3078	1306	1106
16	6	0	0	14722	8125	7837	4460	4002	2099
17	39	111	663	4028	7085	10991	4261	2614	7391
18	23	61	143	5533	1078	1157	918	1544	0
19	18	877	1021	82	1636	4926	1136	1373	1772
20	5	1257	567	5819	1556	3178	3252	1549	63
21	49	2349	12	5422	3028	14098	2352	2794	1503
22	0	180	2121	2939	3854	7702	112	6203	69615
23	1	4081	232	993	1598	3085	813	1113	42894
24	0	814	67	10408	1068	2185	2371	11577	31773
25	10	11524	632	5742	4091	4017	1917	2846	22200
26	142	45	1162	565	2881	9513	1912	995	1
27	40	54	188	1213	3607	7091	1465	83	2630
28	1	151	455	30169	4336	4539	7181	4228	3230
29	85	719	8330	1175	3924	11496	1497	805	2212
30	21	2304	2029	2700	3484	3767	5139	2096	717
31	88	548	2711	0	3530	7267	1032	1714	15033
32	43	812	5459	5804	5900	13050	9367	601	4256
33	6	1683	3922	5800	1864	2678	1304	551	1626
34	5	340	1139	2876	13157	23079	3473	6	167
35	43	266	10912	2059	11671	24806	617	4213	307
36	8	881	7342	2	2981	4493	4569	4327	133
37	1	650	1495	0	9839	21625	1210	648	6497
38	0	0	380	1940	3426	10968	0	0	76
39	61	399	817	0	1132	2570	407	0	0
40	260	1361	1911	2	3928	13973	6151	0	884

most participants. The overall uncertainty should, of course, include information on bounds for model error and systematic error.

It is useful to define at the outset how the terms 'regression' and 'classical factor analysis' are to be understood in this report. A 'regression' model of source apportionment is the same as a chemical mass balance (CMB) model except that the former includes an additive intercept constant determined along with the source intensities in the fitting procedure. 'Classical factor analysis' generally means initial factor extraction by principal component analysis of the correlation matrix of elemental concentration pairs, retention of a number of factors equal to the number of correlation matrix eigenvalues greater than unity, and finally orthogonal VARIMAX rotation. These choices are generally consistent with the default options present in most factor analysis computing packages.

(1) *Effective variance CMB* [*CMB*(*EV*-1)] (SLH)*

(a) Include all source profiles and elements in the initial CMB fit.

(b) Remove sources with negative predicted intensities one at a time, discarding first the source where σ_j/S_j is largest, and repeat the CMB calculation after each removal.

(c) When all remaining source intensities are positive, remove any additional sources one at a time, discarding first the source where σ_j/S_j is largest and greater than 1, and repeat the CMB calculation after each removal.

---

*The descriptions of data evaluation methods 1–7 were provided by the authors whose initials are shown in parentheses.

Table 5B. True mass contributions (ng m^{-3}) by source type and observation period for Set II

Period	Steel-A	Steel-B	Oil-A	Incin-erator	Oil-B	Coal-A	Aggre-gate	Glass	Basalt	Soil	Wood	Coal-B	Road
1	243	27	1903	4667	112	1651	84	723	4469	2236	275	935	1535
2	1546	338	190	477	247	2763	977	1147	11549	9561	2972	2246	5935
3	292	0	68	2648	495	6210	308	1016	7845	5688	486	0	5269
4	1566	292	206	266	479	1117	585	4659	11598	10756	2773	126	6082
5	127	0	99	16563	434	2360	1366	1247	9934	4960	685	0	5263
6	51	5	280	772	612	16484	273	275	22385	10235	698	0	3277
7	607	0	878	661	692	2354	276	49	34763	14038	259	0	1773
8	1668	0	442	582	150	98	353	3	21112	14094	454	3314	263
9	88	391	42	185	67	806	104	31	4547	867	139	395	221
10	133	973	42	0	217	2141	644	421	4321	1798	1764	1120	3745
11	33	770	1128	1887	33	0	1150	4	6968	4714	229	1090	2189
12	509	207	83	2	1899	28	2162	1040	17189	25224	2402	1484	5255
13	18	0	460	834	0	1099	0	277	14794	20442	343	6346	2560
14	116	0	426	680	123	938	172	43	16476	13609	607	436	890
15	632	0	669	205	214	719	255	0	9890	3125	345	0	1037
16	1042	0	329	1993	194	4202	918	0	16970	22546	896	0	843
17	254	0	41	9022	368	1459	340	2	9806	11302	366	2724	2951
18	753	0	404	1546	216	1182	413	0	5233	1673	92	0	0
19	238	0	602	508	404	0	288	1136	5492	1198	401	1874	2075
20	95	0	297	273	409	1734	576	215	6902	2747	690	4	927
21	394	93	1990	389	318	0	243	67	9141	8259	288	10167	2075
22	77	1176	367	2878	200	5740	78	148	9291	2483	2645	12906	9897
23	55	790	155	4238	103	2025	115	587	3388	2129	588	4873	9072
24	357	363	264	3955	308	129	33	79	3980	181	1301	8629	10644
25	64	483	411	6153	243	0	11	0	3504	4070	411	1085	5806
26	575	379	2721	60	2237	43515	539	753	10382	3784	3589	1856	3436
27	53	3	1921	439	809	13603	111	625	10322	8648	395	829	2736
28	498	30	268	3396	823	8086	1212	106	14646	5812	766	1013	2105
29	911	0	349	99	799	4958	472	405	15910	4973	300	1938	2263
30	614	307	200	148	648	0	988	519	7598	5377	1568	4343	3195
31	87	979	432	0	126	1743	1172	382	13264	8255	347	1503	2388
32	251	2493	2131	219	209	0	1397	125	13533	3952	1163	8806	3230
33	94	315	453	1684	24	434	81	74	3338	2289	360	6183	2620
34	135	256	777	92	214	3969	86	86	27525	8207	1360	78	7487
35	1508	0	280	199	0	0	0	819	10453	6489	713	27368	2889
36	24	0	83	1560	394	0	3	83	9674	438	739	110	1849
37	12	0	15	2	149	0	232	207	15127	2151	496	4335	1526
38	110	205	169	669	262	898	111	336	8802	797	1120	12466	2657
39	443	247	576	145	481	3847	42	219	3570	2031	258	0	0
40	607	388	239	2390	688	10918	301	666	17465	19755	2629	0	5833

Table 5C. True mass contributions (ng m^{-3}) by source type and observation period for Set III

Period	Steel-A	Steel-B	Oil-A	Incin-erator	Oil-B	Coal-A	Aggre-gate	Glass	Basalt	Soil	Wood	Coal-B	Road
1	247	24	1991	5087	122	1610	81	708	4432	2217	295	1032	1485
2	1633	366	185	507	265	2591	983	1113	11550	10005	2696	2396	6050
3	343	0	70	2788	531	6146	296	999	8171	5768	584	0	4745
4	1671	317	221	264	517	1180	594	4633	11717	10650	2566	202	3786
5	126	0	92	72341	434	2216	1328	1261	9807	4492	605	0	4707
6	52	5	270	860	630	18077	280	272	22100	10154	840	0	3183
7	593	0	906	669	747	2302	274	54	35417	14320	265	0	1702
8	1676	0	442	1235	155	102	341	3	21814	14910	388	3280	261
9	99	383	44	361	74	824	104	33	4716	847	131	585	207
10	135	1118	41	0	214	2203	670	459	4242	1680	1828	1489	3566
11	28	756	1127	1876	32	0	1088	4	7547	4880	255	962	2195
12	488	218	79	2	1841	29	2257	1232	17818	25499	3225	1451	6067
13	20	0	439	1395	0	1100	0	295	16361	19748	289	10309	2785
14	98	0	425	638	139	891	181	42	16957	14726	559	396	860
15	645	0	718	490	222	771	251	0	10123	3176	339	0	1035
16	1003	0	372	2164	199	4046	961	0	17124	22406	316	0	985
17	313	0	43	7565	353	1530	334	2	10160	11450	343	2592	2773
18	782	0	422	1443	219	1165	436	0	5715	1622	79	0	0
19	258	0	580	525	403	0	296	1227	5484	1254	498	2278	1903
20	97	0	286	285	413	1690	563	219	6893	2967	604	5	827
21	464	85	1791	443	321	0	257	68	9345	9036	272	10498	2039
22	94	1246	363	3899	194	5401	79	134	9447	2583	2035	13975	10618
23	57	1032	157	5325	100	2215	108	598	3624	2112	657	5307	9918
24	385	385	285	3894	321	140	34	75	3961	173	1558	9394	9685
25	78	556	387	4510	227	0	12	0	3713	3914	458	937	6886
26	665	388	2849	337	2473	46760	544	799	10853	4159	4233	2010	3833
27	53	3	1910	373	870	13716	109	677	10538	9874	362	696	2667
28	581	29	287	4373	970	7857	1252	103	15326	5902	617	991	2205
29	826	0	336	110	836	5134	484	411	15773	4763	257	1911	2076
30	566	315	197	132	625	0	956	497	7257	5257	1373	4663	3529
31	105	1012	457	0	140	1856	1202	398	13755	8135	383	1799	2382
32	268	2379	2069	342	229	0	1407	128	13705	4018	1121	8114	2990
33	88	346	463	2029	25	439	82	79	3276	2278	473	6576	3209
34	127	293	769	119	217	4139	85	82	27302	8150	1531	76	7750
35	1608	0	261	249	0	0	0	836	10120	6632	675	35466	3308
36	30	0	85	1907	430	0	3	84	9766	460	762	106	1983
37	18	0	16	2	161	0	228	203	15635	2049	524	4757	1845
38	117	193	171	716	288	892	107	361	9199	818	1024	13923	2384
39	473	250	557	137	505	3928	44	234	3580	1999	313	0	0
40	685	446	222	4609	850	10720	312	695	18333	22929	3153	0	7058

Table 6. Emissions inventory for Sets II and III

Source	Emission rate ($t\ y^{-1}$)
Steel-A	70,000
Steel-B	107,000
Oil-A	43,000
Incinerator	47,000
Oil-B	49,000
Coal-A	503,000
Aggregate	31,600
Glass	48,400
Basalt	50,000
Soil	862,000
Auto	2,285,000
Wood	934
Coal-B	795,000

(d) Experiment with adding or substituting discarded sources to improve agreement between the measured and predicted elemental concentrations to reduce the percentage error in predicted total mass, and to minimize the reduced chi-square value.

This procedure was applied to the individual observation periods separately and also to the average of the periods. Although much more computing is required, the former was considered preferable because of the more satisfactory handling of multicollinearity: even though a particular source is discarded in the analysis of one observation period, it will likely be retained in another, and average intensities for all sources are then obtained from an average of results from all periods.

Uncertainties in the source contributions were obtained differently in the two treatments. When the procedure was applied to the average of the individual periods, the uncertainties in the average concentrations were estimated by propagating the errors (Bevington, 1969) for the individual periods through the averaging process

$$\sigma_{\bar{C}_i} = \frac{1}{N}\left(\sum_{K=1}^{N} \sigma^2_{C_i}(K)\right)^{1/2}, \tag{2}$$

where $N = 40$. The CMB procedure was then applied using the average concentrations and their uncertainties, and the uncertainties in the average source contributions were the uncertainties estimated by the effective variance method.

When the procedure was applied to the individual periods, the uncertainties in the average source contributions were estimated by propagating through the averaging process the uncertainties estimated by the effective variance method for each period

$$\sigma_{\bar{S}_j} = \frac{1}{N}\left(\sum_{K=1}^{N} \sigma^2_{S_j}(K)\right)^{1/2}. \tag{3}$$

When a source was not included in the CMB for a particular period, the uncertainty in its zero contribution was assumed also to be zero. This assumption is probably not completely valid, but a means to estimate the 'detection limit' for sources excluded from the CMB has not yet been developed.

The uncertainty estimates obtained via Equation (3) using the individual period results were substantially smaller than those from the average ambient composition approach. As will be seen in Table 13, many of the estimated source contributions obtained by the former method deviate from the 'true' contributions by more than would be expected from the uncertainty estimates. This suggests that the method used to estimate the uncertainties in results from individual periods may not be appropriate.

(2) *Effective variance CMB* [*CMB(EV-2)*] (JAC)

This procedure is the same as the preceding one. It was applied only to the individual observation periods separately and the results averaged. In place of Equation (3), however, uncertainties in the average source intensities were calculated from

$$\sigma_{\bar{S}_j} = \left(\sum_{K=1}^{N} \frac{(S_j^{(K)} - \bar{S}_j)^2}{N(N-1)} w\right)^{1/2}. \tag{4}$$

The CMB(EV-2) results were submitted after completion of the workshop, when the true source intensities were known.

(3) *Weighted ridge regression CMB* [*CMB(RR)*] (HJW)

(a) Calculate variance inflation factors (VIFs) for each source profile (Henry *et al.*, 1984).

(b) Perform weighted ridge regression for a series of ridge *k* values between 0 and 0.72.

(c) Select the solution where (1) the calculated regression coefficients (source intensities and intercept) do not change appreciably for a small change in *k*, (2) no significant negative coefficients are calculated and (3) the calculated total mass is in good agreement with the measured total mass.

The VIFs of Step (a) are not required in the subsequent steps, but serve as a diagnostic in establishing the need for ridge regression. Twenty-eight *k* values were used, more closely spaced near 0 than near 0.72. Like CMB(EV), the procedure can be applied both to individual observation periods and to the average of the periods. The latter was considered preferable because it presents less of a problem for calculating valid uncertainty values and requires fewer computing resources.

(4) *Weighted CMB/factor analysis* [*CMB(W)/FA*] (GDT)

(a) Perform classical factor analysis on all *elemental observations* to screen data and determine number and characteristics of major sources.

(b) Perform classical factor analysis on all available *source profiles* to classify sources into distinct source groups.

(c) Attempt to match, based on correspondences in elements with heavy loadings, each ambient factor from Step (a) with a source group factor from Step (b).

(d) For each source group factor that has been matched, choose one source profile as representative of it and use the resulting set of source profiles in a CMB calculation weighted by the ambient concentration uncertainties only.

Step (a) applied to Set I initially identified (on the basis of the number of eigenvalues exceeding unity) 6 major factors, as shown in Table 7. One factor, however, consisted almost entirely of arsenic. Since this factor matched none of the source profiles and since arsenic was a very 'noisy' variable (mean concentration < uncertainty), arsenic was eliminated from further consideration, thus reducing the solution to five major factors. Table 8 gives the results of Step (b), showing seven source groups: (1) crustal: *soil*, basalt, coal-A, aggregate furnace; (2) carbon-rich: *automotive*, wood combustion; (3) glass manufacturing; (4) *incinerator*; (5) steel furnaces: steel-A, *steel-B*; (6) coal-B; (7) oil combustion: *oil-A*, oil-B. Four of the five ambient factors could be reasonably matched to the first four source groups listed above. This was accomplished by comparing the factors listed in Table 7 with the supplied elemental profiles of the sources in the listed groups. The lack of a good match for the remaining ambient factor (Factor 3) was taken to indicate (correctly) that an important source profile

Table 7. Factor analysis of Set I by CMB(W)/FA*

	Rotated factor pattern					
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
C	0.26	0.92	0	0	0	0
Na	0.30	0.79	0	0.44	0	0
Al	0.92	0.33	0	0	0	0
Si	0.84	0	0.39	0	0	0
S	0	0	0	0.58	0	0.73
Cl	0	0	0	0.95	0	0
K	0	0	0	0.96	0	0
Ca	0.88	0	0	0	0	0
Ti	0.97	0	0	0	0	0
V	0.59	0	0	0	0	0.74
Cr	0	0	0.97	0	0	0
Mn	0.89	0.35	0	0	0	0
Fe	0.92	0	0.26	0	0	0
Ni	0	0	0.97	0	0	0
Cu	0	0	0.96	0	0	0
Zn	0	0.43	0.28	0.83	0	0
As	0	0	0	0	0.97	0
Br	0	0.96	0	0	0	0
Pb	0	0.95	0	0	0	0
Percentage of total variance explained by each factor						
	30.3	20.8	16.9	17.4	5.3	6.8

* Loadings of < 0.25 have been replaced by 0.

Table 8. Factor analysis of source profiles by CMB(W)/FA*

	Rotated factor pattern						
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
Steel -A	0	0.45	0	0	0.88	0	0
Steel-B	0	0.54	0	0	0.83	0	0
Oil-A	0.75	0	0.40	0	0	0	0.47
Incin	0	0	0	0.97	0	0	0
Oil-B	0	0.79	0.31	0	0.27	0	0.33
Coal-A	0.97	0	0	0	0	0	0
Aggregate	0.99	0	0	0	0	0	0
Glass manuf.	0	0	0.97	0	0	0	0
Basalt	0.95	0	0	0	0	0	0
Soil	0.97	0	0	0	0	0	0
Auto	0	0.94	0	0	0.25	0	0
Wood	0	0.93	0	0	0.31	0	0
Coal-B	0.72	0	0.25	0	0	0.62	0
Percentage of total variance explained by each factor							
	38.2	23.2	10.2	7.8	13.5	3.5	3.0

* Loadings of < 0.25 have been replaced by 0.

had been omitted from the world list. For lack of a better alternative, the source profile steel-B was chosen as a surrogate for the missing source, on the basis of its high loadings of chromium, nickel and copper. The five representative source profiles chosen for Step (d) are italicized in the above. Step (d) was applied to each observation period, and the source intensities obtained for the individual periods averaged. The uncertainties in the average source intensity estimates were calculated by Equation (3), with the difference that N counted only the period for which source j was used in a CMB (i.e. $S_j^{(K)} \neq 0$).

Step (a) was repeated with several meteorological variables added to the elemental ones: percentage of wind direction (NE, SE, SW, NW), percentage of stable conditions, percentage of unstable conditions, and mixing height during the sampling period. The 4-factor solution for Set I is shown in Table 9. The most conspicuous feature is the association of NE wind direction with carbon-rich Factor 2. This was taken as indicative of a major source of automotive emissions to the northeast, a conclusion that appears consistent with the automotive corridor in the northeast sector of City Plan 1 (Fig. 1).

(5) *Effective variance CMB/factor analysis* [*CMB(EV)/FA*] (CWL)

(a) Perform classical factor analysis on all elemental observations to screen data and determine number and characteristics of major sources.

Table 9. Factor analysis of Set I including meteorological variables by CMB(W)/FA*

	Rotated factor pattern			
	Factor 1	Factor 2	Factor 3	Factor 4
C	0.36	0.86	0	0
Na	0.36	0.76	0	0.47
Al	0.91	0.29	0	0
Si	0.86	0	0.40	0
S	0.36	0	0	0.77
Cl	0	0	0	0.91
K	0	0	0	0.91
Ca	0.90	0	0	0
Ti	0.96	0	0	0
V	0.77	0	0	0.35
Cr	0	0	0.95	0
Mn	0.88	0.31	0	0
Fe	0.93	0	0.28	0
Ni	0	0	0.95	0
Cu	0	0	0.94	0
Zn	0	0.44	0.30	0.80
Br	0.26	0.92	0	0
Pb	0.25	0.91	0	0
PERCNE	0	0.84	0	0
PERCSE	0	0	0.46	0.40
PERCSW	0.36	−0.43	0	0
PERCNW	0	0	−0.33	0.29
PERCUNST	0	0.27	0	0
PERCSTBL	0.46	0	0	0.29
MIXHT	0	0	0.44	0
Percentage of total variance explained by each factor				
	26.4	18.4	14.9	15.4

* Loadings of < 0.25 have been replaced by 0.

(b) On the basis of similar prominent elements, associate a single representative source profile with each factor found in Step (a).

(c) Perform a CMB(EV) with the smaller set of source profiles chosen in Step (b).

This procedure is essentially a simpler but less systematic and less objective version of CMB(W)/FA. Step (a) applied to Set I produced results virtually identical to those produced by CMB(W)/FA, with five nontrivial factors identified (arsenic excluded). In Step (b), after some trial and error, soil, auto, incinerator, oil-B and steel-A were selected as representative of the five factors. In Step (c), the effective variance CMB program of Dzubay *et al.* (1981) was applied only to the average of the Set I observations, using all elemental concentrations except arsenic.

Step (a) was repeated, adding the meteorological variables percentage of the wind direction and average wind speed. In contrast to CMB(W)/FA, none of the new variables was found to be significantly associated with the old factors. The reason for this difference is not understood.

(6) *Multiple linear regression by tracers/factor analysis* [*MLR(T)/FA*] (SLD)

(a) Perform classical factor analysis on all elemental observations to screen data and determine number and characteristics of major sources.

(b) For each factor found in Step (a), choose the element with the highest loading as the tracer for that factor.

(c) Perform a backward stepwise unweighted multiple linear regression with total mass as the dependent variable and the tracer elements as independent variables. Remove an independent variable whenever its regression coefficient is < 1.96 times the calculated uncertainty in the coefficient.

(d) Calculate the intensity of source j during observation period K, $S_j^{(K)}$, from

$$S_j^{(K)} = B_j C_j^{(K)}, \qquad (5)$$

where B_j is the regression coefficient found in Step (c) for the tracer element of source j, and $C_j^{(K)}$ is the observed concentration of that tracer for observation period K.

Step (a) applied to Set I produced the same results as the corresponding step in the two previous procedures. Step (b) determined titanium, bromine, nickel, potassium and vanadium as the tracer elements for the five nontrivial sources found in Step (a). The estimated source intensities obtained in Step (d) were averaged across observation periods. The uncertainty in each average was obtained in a relatively crude way, by only considering the standard error of B_j.

A feature of MLR(T)/FA is its ability to generate its own source profiles from the ambient observations, without requiring the input of any external source profiles. Source profiles were generated by repeating Step (c) with each element, in succession (instead of just total mass), as the dependent variable. (The cases involving a tracer element as the dependent variable are trivial, resulting in one regression coefficient of unity and the rest of 0.) The matrix of regression coefficients thus obtained was then converted to the normal form of a source profile matrix by applying a normalization constant to each column (i.e. source) such that the total mass regression coefficient became unity. The resulting source profile matrix for Set I is given in Table 10A. The true profiles are given in Table 10D.

(7) *Target transformation factor analysis* [*TTFA*] (PKH)

(a) Perform 'classical' factor analysis on all elemental observations to screen data and determine an upper limit on the number of major sources.

(b) Perform target transformation factor analysis to determine the number of major sources, their average intensities, and their profiles.

The 'classical' analysis in Step (a) was based on correlations between observation periods (Q-mode analysis), whereas the three preceding factor analyses used correlations between elemental concentrations (R-mode). The upper limit in Step (a) was determined by the number of factors that accounted for at least one unit of variance after orthogonal (VARIMAX) rotation. This also differs from the three previous procedures, which used the more common method of equating the number of significant factors to the number of correlation matrix eigenvalues greater than unity. For Set I (after eliminating arsenic), Step (a) indicated a maximum of seven factors, compared with the five found in the three previous procedures. In Step

Table 10A. Source profiles (mg g^{-1}) for Set I from MLR(T)/FA

Element	Soil	(Road)	(Sandblast)	Incinerator	Oil
C	NS*	637 ± 31	NS	NS	240 ± 57
Na	NS	62.2 ± 4	17.5 ± 6	94.2 ± 12	22.2 ± 7
Al	245 ± 17	56.5 + 12	NS	NS	NS
Si	652 ± 47	NS	363 ± 58	NS	NS
S	−40.3 ± 13	NS	NS	168 ± 22	135 ± 16
Cl	−22.3 ± 4	11.0 ± 2	25.4 ± 4	277 ± 8	NS
K	NS	NS	NS	192	NS
Ca	53.5 ± 6	7.84 ± 3	NS	NS	18.8 ± 7
Ti	19.8	NS	NS	NS	NS
V	NS	NS	NS	NS	1.06
Cr	NS	NS	6.92 ± 0.2	NS	NS
Mn	2.8 ± 0.2	0.735 ± 0.1	0.679 ± 0.2	NS	NS
Fe	144 ± 5	NS	46.1 ± 6	NS	NS
Ni	NS	NS	4.12	NS	NS
Cu	NS	NS	7.65 ± 0.29	2.50 ± 0.8	NS
Zn	NS	3.58 ± 0.3	5.55 ± 0.5	26.5 ± 1	NS
As	NS	NS	NS	NS	NS
Br	NS	49.0	NS	NS	NS
Pb	NS	140 ± 2	NS	NS	NS

* NS = no solution (no non zero ratio found), i.e. $\beta_i < 1.96\,\sigma_{\beta_i}$.

Table 10B. Source profiles (mg g^{-1}) for Set I from TTFA

Element	Wood	Soil	Incinerator	Oil	(Sandblast)	(Road)
C	464 ± 24	34 ± 10	33 ± 40	140 ± 40	3.6 ± 12.8	642 ± 30
Na	8.3 ± 9.0	4.3 ± 3.8	115 ± 16	12 ± 12	13 ± 5	84.7 ± 11.2
Al	0 ± 50	256 ± 21	76 ± 81	8.0 ± 63.1	83 ± 27	9.1 ± 59.6
Si	15 ± 80	518 ± 34	0.001 ± 0.646	37 ± 105	375 ± 44	11 ± 99
S	0.040 ± 0.123	12 ± 6	149 ± 23	87 ± 18	11 ± 8	47 ± 17
Cl	3.7 ± 5.1	4.2 ± 2.2	245 ± 9	6.2 ± 6.7	2.2 ± 2.8	16 ± 7
K	0.26 ± 7.96	9.6 ± 3.4	213 ± 14	7.9 ± 10.6	17 ± 5	0.51 ± 9.94
Ca	0 ± 18	7.3 ± 7.6	12 ± 31	41 ± 24	49 ± 10	54 ± 23
Ti	0.41 ± 2.43	11.3 ± 1.1	0.85 ± 4.11	4.6 ± 3.3	3.7 ± 1.4	8.5 ± 3.1
V	0.074 ± 0.216	0.30 ± 0.10	0.052 ± 0.364	1.94 ± 0.29	0.14 ± 0.12	0.007 ± 0.688
Cr	0.25 ± 0.49	0.78 ± 0.21	0.088 ± 0.821	0.11 ± 0.65	3.58 ± 0.27	0.78 ± 0.61
Mn	0.42 ± 0.33	2.20 ± 0.14	0.22 ± 0.55	0.36 ± 0.43	0.83 ± 0.18	1.1 ± 0.5
Fe	0.69 ± 3.38	85.0 ± 5.7	36 ± 23	4.4 ± 17.7	113 ± 8	69 ± 16.8
Ni	0.080 ± 0.330	0.29 ± 0.14	0.50 ± 0.56	0.93 ± 0.44	2.09 ± 0.18	0.71 ± 0.42
Cu	0.001 ± 0.398	0.18 ± 0.23	6.12 ± 0.92	0.63 ± 0.72	4.18 ± 0.30	0.98 ± 0.68
Zn	0.42 ± 1.66	1.2 ± 0.7	12 ± 3	0.52 ± 2.19	3.1 ± 1.0	5.4 ± 2.1
Br	0.014 ± 0.952	0.009 ± 0.400	0.021 ± 0.609	0.037 ± 0.257	0.010 ± 0.519	73.3 ± 1.2
Pb	0.27 ± 3.04	1.5 ± 1.3	13 ± 6	3.4 ± 4.1	0.35 ± 1.66	223 ± 4

Table 10C. Source profiles (mg g^{-1}) for Set I from TTFA*

Element	Wood	Soil	Incinerator	Oil	Coal	Sandblast	Road
C	591±42	68±12	35±37	185±26	15±51	16±25	573±17
Na	4.0±20	0.0061±5	64±18	14±13	35±25	8.5±11.8	54.8±8
Al	10±31	166±9	0.0002±26	11±20	210±40	0.00099±18	44±13
Si	12±92	466±25	0.53±79	44±57	230±120	508±54	73±37
S	0.25±26	5.7±7.2	64±24	120±17	180±40	0.025±15	10±11
Cl	16±11	0.35±2.90	196±10	5.4±6.7	23±14	37±7	16±5
K	13±5	7.5±1.3	122±4	2.9±2.8	17±6	8.5±2.7	0.00004±1.7
Ca	14±14	41.4±3.6	1.6±11.4	26±9	81±16	1.5±7.7	13±6
Ti	0±3	12.9±0.7	5.4±2	2.7±1.4	16±3	2.7±1.4	0.77±0.90
V	0±0.11	0.315±0.030	0.25±0.10	0.660±0.068	0.41±0.14	0.15±0.07	0.00029±0.043
Cr	0±0.4	0.54±0.09	1.2±0.3	0.026±0.203	0.35±0.40	5.49±0.20	0.027±0.130
Mn	0±0.4	1.97±0.09	1.3±0.3	0.13±0.20	1.9±0.4	1.1±0.2	0.61±0.13
Fe	0.31±9.32	96.4±2.6	2.4±8.2	35±6	143±12	45.6±5.5	9.2±3.8
Ni	0±0.11	0.196±0.029	0.12±0.10	0.094±0.066	0.50±0.13	3.06±0.07	0.0037±0.042
Cu	0±0.7	0.52±0.19	4.13±0.59	0.40±0.42	0.014±0.820	6.63±0.40	0.060±0.269
Zn	0.45±2.43	0.28±0.66	10±3	0.25±1.52	0.53±2.97	6.0±1.5	4.2±1.0
Br	0.12±1.13	0.89±0.31	0.37±0.98	0.48±0.71	0±1.4	0.70±0.67	52.4±0.5
Pb	0.014±4.61	5.0±1.3	9.2±4.1	3.4±2.9	1.6±5.7	5.9±2.8	152±2

* Submitted after completion of workshop.

Table 10D. True source profiles (mg g^{-1}) for Set I*

Element	Oil-B	Incinerator	Soil	Road†	Wood	Sandblast	Coal-B
C	200	40	66.4	429	600	5.0	8.0
Na	22	120	4.0	52.5	3.0	8.0	4.0
Al	5.4	4.2	222	62	7.0	7.5	140
Si	70	7.0	543	147	8.9	400	200
S	110	130	0.26	6.06	1.00	8.8	180
Cl	12	270	0.13	8.06	10.0	28	17
K	1.40	200	12.0	5.0	8.0	7.0	8.0
Ca	8.6	11.0	20.5	13.9	9.0	8.2	40
Ti	1.10	0.90	12.7	3.17	0.020	4.0	9.0
V	0.67	0.027	0.25	0.143	0	0.031	0.8
Cr	0.42	0.85	0.31	0.141	0.002	5.8	0.3
Mn	0.140	0.33	2.0	0.94	0.020	1.4	0.2
Fe	13.0	6.1	97	29.5	1.00	63	70
Ni	0.06	0.100	0.090	0.070	0.005	3.2	0.6
Cu	0.75	3.6	0.20	0.050	0	6.9	0.4
Zn	1.30	26	0.41	2.51	0.70	4.7	0.9
As	0.25	0.150	0.010	0.010	0	0.099	0.2
Br	0.39	0.82	0.010	31.3	0.080	0.047	0.01
Pb	3.8	17.0	0.26	90.1	0.100	5.8	1.2

* Steel-A and basalt sources were also present but were not deduced by any workshop participants.

† Road is a linear combination of soil (25%) and auto (75%).

(b), several indicator parameters (Table 11A) were examined to arrive at the final number of source factors. The chi-square and average error indicators were the most useful in fixing this number at 6. (An indicator parameter should show a smaller rate of decrease once the proper number of factors has been used.)

Details of the procedure used in Step (b) are given by Roscoe and Hopke (1981a). Like MLR(T) FA, TTFA generates its own source profiles without requiring the input of any source information. The source profiles resulting from TTFA applied to Set I are given in Table 10B, to be compared with the true profiles in Table 10D. The uncertainties given in Table 10B represent an upper limit for one standard deviation, as discussed by Roscoe and Hopke (1981b).

After the workshop concluded, a second TTFA Set I solution was submitted that is regarded as superior to the first one. For the second solution, the chi-square and average error indicators (Table 11B) showed a 7-factor solution, in agreement with the classical factor analysis of Set I. The source profiles generated in the second solution are given in Table 10C. The second solution used an R-mode analysis instead of the Q-mode employed in the first solution.

RESULTS OF THE INTERCOMPARISON

Tables 12 and 13, which give the average source strengths determined by the Quail Roost II participants from their analyses of simulated Data Sets I and

Reproduced from
best available copy.

Table 11A. Diagonalization results and reproduction summary for Set I from TTFA*

Factor	Eigenvalue	r.m.s. error	Chi-square	Exner	Malinowski empirical indicator	Average error (%)
1	3.60E+01	4.48E+01	3.20E+02	0.346	4.75E−05	130
2	2.97E+00	1.48E+01	8.11E+01	0.152	2.23E−05	70
3	3.26E−01	1.09E+01	6.30E+01	0.112	1.75E−05	58
4	2.37E−01	7.30E+00	6.94E+01	0.071	1.18E−05	51
5	8.10E−02	4.65E+00	8.18E+01	0.048	8.75E−06	48
6	3.19E−02	3.06E+00	1.78E+01	0.036	7.06E−06	37
7	2.18E−02	2.14E+00	1.55E+01	0.025	5.26E−06	34
8	1.25E−02	1.59E+00	9.28E+00	0.0156	3.52E−06	25
9	5.65E−03	7.19E−01	1.04E+01	0.0078	1.90E−06	23
10	1.42E−03	4.57E−01	5.62E+00	0.0038	1.02E−06	14
11	2.66E−04	3.38E−01	2.80E+00	0.0025	7.35E−07	10
12	9.80E−05	2.55E−01	5.51E+00	0.0018	5.72E−07	10
13	6.92E−05	1.17E−01	1.70E+00	0.0010	3.50E−07	6.4
14	1.50E−05	6.51E−02	1.62E+00	0.0007	2.75E−07	5.1
15	9.50E−06	4.14E−02	1.43E+00	0.0004	1.95E−07	3.7
16	2.95E−06	3.14E−02	1.51E+00	0.0003	1.63E−07	2.9
17	2.29E−06	1.94E−02	1.68E+00	0.0002	1.17E−07	1.9

* Six-factor solution submitted prior to workshop.

Table 11B. Diagonalization results and reproduction summary for Set I from TTFA*

Factor	Eigenvalue	r.m.s. error	Chi-square	Exner	Malinowski empirical indicator	Average error (%)
1	1.35E+01	4.9E+01	9.5E+02	0.75	2.80E−04	180
2	1.92E+00	3.4E+01	4.0E+02	0.56	2.46E−04	115
3	1.12E+00	2.5E+01	4.0E+02	0.42	2.16E−04	97
4	1.01E+00	1.8E+01	1.3E+01	0.22	1.39E−04	29
5	2.48E−01	1.6E+01	1.1E+01	0.14	1.06E−04	23
6	4.46E−02	1.3E+01	1.3E+01	0.12	1.11E−04	23
7	3.17E−02	1.1E+01	3.1E+00	0.107	1.18E−04	14
8	2.46E−02	8.3E+00	2.2E+00	0.091	1.28E−04	12
9	1.67E−02	5.8E+00	2.7E+00	0.079	1.45E−04	11
10	1.08E−02	3.5E+00	3.3E+00	0.070	1.72E−04	11
11	9.53E−03	3.2E+00	3.4E+00	0.060	2.09E−04	10
12	9.24E−03	2.5E+00	4.3E+00	0.050	2.54E−04	8.9
13	8.42E−03	1.5E+00	3.9E+00	0.038	3.06E−04	6.6
14	5.15E−03	1.2E+00	3.4E+00	0.028	4.01E−04	4.7
15	3.18E−03	9.5E−01	4.2E+00	0.020	5.93E−04	3.5
16	1.44E−03	7.4E−01	7.7E+00	0.0157	1.24E−03	3.3
17	1.31E−03	3.8E−01	1.7E+01	0.0091	4.02E−03	3.0

* Seven-factor solution submitted after completion of workshop.

II, are the most important in this report. No results for Set III were reported.

In presenting the results in an understandable and equitable manner, two complications must be addressed. The first relates to the difference between the CMB and factor analysis approaches to source apportionment. The former uses a supplied collection of source profiles, while the latter generates its own from the variability of the ambient data, without prior reference to the supplied profiles. In general, the two groups of profiles will not agree exactly, and judgment is needed to make a correspondence between them. In the case of Data Set I, factor analysis was part of the approach used by every reporting participant, and there was fairly good agreement on the number and nature of the sources found (see Tables 10A–D). Accordingly, for Set I the supplied sources were grouped in a manner to best correspond to the source categories identified in the factor analyses. While this step was necessarily somewhat arbitrary, its basis was the information in Tables 10A–D and Table 8. The groupings are indicated in Table 12. No such problem arises with Data Set II, since only CMB analyses were reported for that set.

The second complication arises from the use of the road source (a fixed linear combination of the supplied auto and soil sources) in the generation of all three data sets. A factor analysis approach is by nature sensitive to

Table 12. Estimated source impacts ($\mu g\ m^{-3}$) compared with true values for Set I*

Source	CMB(EV)/FA†	CMB(W)/FA‡	MLR(T)/FA§	TTFA	TTFA‖	Truth
Steel-A	1.0 ± 0.3	—	—	—	—	0.05
Steel-B		—	—	—	—	—
Oil-A	6.1 ± 1.0	8.7 ± 0.9	7.9 ± 0.8	2.3 ± 0.4	2.1 ± 1.0	—
Oil-B						2.0
Incinerator	1.5 ± 0.2	0.7 ± 0.3	2.4 ± 0.4	1.8 ± 0.1	1.9 ± 0.14	1.3
Glass	—	—	—	—	—	—
Coal-B	—	—	—	—	2.2 ± 0.73	2.4
Coal-A	15.9 ± 1.1	15.9 ± 0.2	10.6 ± 0.9	12.7 ± 0.5	12.5 ± 0.75	—
Aggregate						—
Basalt						4.7
Soil¶						8.0**
Road¶	5.3 ± 0.4	8.3 ± 0.2	5.0 ± 0.2	3.0 ± 0.1	4.0 ± 0.09	7.1
Wood	—	—	—	7.4 ± 0.7	4.3 ± 0.52	3.3
Sandblast††	—	1.9 ± 0.4	4.0 ± 0.3	5.2 ± 0.4	4.1 ± 0.20	4.2
Total	29.8	35.5	29.9	32.4	31.1	33.05

* Uncertainties represent one standard deviation.

† CMB result for the average of the observation periods.

‡ Average of CMB results for each observation period.

§ Fitted intercept = 3.3 ± 1.5 $\mu g\ m^{-3}$.

‖ Results submitted after completion of workshop.

¶ Data were generated using road as a source component, where road = 0.25 soil + 0.75 auto.

** Soil from all sources (including road component) = 9.8 $\mu g\ m^{-3}$.

†† Sandblast was an extraneous source, not included among the 13 profiles furnished to the participants. Labels originally used for this source by those who furnished estimates were: 'other (steel-B used as surrogate)', CMB(W)/FA; 'unknown (nickel tracer)', MLR(T)/FA; and 'aggregate', TTFA.

Table 13. Estimated source impacts ($\mu g\ m^{-3}$) compared with true values for Set II*

Source	CMB(EV-1)$_1$†	N1‡	CMB(EV-1)$_2$§	CMB(RR)$_1$§	CMB(RR)$_2$§,‖	CMB(EV-2)†,‖,¶	N2**	Truth
Steel-A	0.58 ± 0.033	19	0.97 ± 0.14	0.29 ± 0.73	0.66 ± 2.4	0.55 ± 0.11	21	0.42
Steel-B	0.24 ± 0.017	11	—	0.31 ± 0.49	0.26 ± 1.6	0.22 ± 0.08	8	0.29
Oil-A	0.44 ± 0.033	17	0.47 ± 0.12	0.11 ± 0.66	0.45 ± 0.56	0.47 ± 0.10	21	0.56
Oil-B	1.2 ± 0.13	11	1.7 ± 1.1	2.3 ± 3.3	0.11 ± 2.7	0.72 ± 0.25	8	0.41
Incinerator	1.5 ± 0.05	27	1.5 ± 0.1	1.3 ± 0.84	1.5 ± 0.29	1.5 ± 0.43	26	1.81
Glass	0.16 ± 0.03	3	—	0.15 ± 1.3	0.31 ± 1.1	0.28 ± 0.11	8	0.46
Coal-B	3.7 ± 0.17	21	3.8 ± 1.1	4.3 ± 3.1	4.9 ± 2.3	3.8 ± 0.91	25	3.3
Coal-A	4.3 ± 0.3	11	6.2 ± 1.3	3.0 ± 4.9	4.6 ± 4.7	5.3 ± 1.40	22	3.7
Aggregate	2.1 ± 0.3	11	—	7.3 ± 4.0	2.5 ± 6.9	2.1 ± 0.57	11	0.46
Basalt	8.2 ± 0.5	24	11.2 ± 1.7	9.0 ± 3.4	8.2 ± 6.1	9.5 ± 1.28	31	11.3
Soil	9.5 ± 0.4	33	7.8 ± 1.4	6.3 ± 2.1	8.6 ± 3.5	8.4 ± 1.26	33	7.9
Auto	2.5 ± 0.05	40	2.5 ± 0.1	2.3 ± 0.78	2.5 ± 0.2	2.5 ± 0.31	40	2.5
Wood	0.5 ± 0.1	11	0.6 ± 0.5	1.1 ± 0.38	1.0 ± 0.54	0.91 ± 0.16	26	0.94
Total	34.92		36.74	37.76	35.59	36.25		34.05

* Uncertainties represent one standard deviation.

† Average of CMB results for each observation period (40 total).

‡ Number of periods (of 40 total) for which source was used in CMB(EV-1)$_1$ (separate analysis of each period).

§ CMB result for the average of the observation periods.

‖ Results submitted after completion of workshop.

¶ Average intercept = 0.52 $\mu g\ m^{-3}$.

** Number of periods (of 40 total) for which source was used in CMB(EV-2).

a combined source rather than its components, whereas a CMB approach resolves the (auto, soil) components and the combination goes unrecognized. For this reason it is natural to express the results from factor analyses (Table 12) in terms of the road source, but the results from CMB analyses (Table 13) in terms of the auto source.

Set I results

Examining the Set I results in Table 12, we conclude the following.

(1) There were no false positives (absent sources reported as present).

(2) False negatives (sources missed) were primarily

restricted to those that were 'weak' (large relative standard deviation [RSD]) or highly 'collinear' (correlated) with other sources. This result is especially evident in the approaches that used factor analysis to ascertain source categories and number of sources. For example, of the nine sources present, TTFA found only six initially and seven in the post-workshop solution.

(3) The unknown source, sandblast, was purposely made large (RSD ≪ 1; see Fig. 3). Three out of the four responding participants reported a value for an extra component, each of which appears quantitatively consistent with the truth.

(4) Arsenic was seen as a lone component in all factor analyses. However, every participant correctly judged this value to be measurement-error-related rather than indicative of a true source impact. Consequently, arsenic was omitted in all the final factor analyses of Set I.

(5) Two of the methods, MLR(T)/FA and TTFA, provided independent estimates of the source profiles, a very interesting achievement. A comparison of the statistically significant estimates listed in Tables 10A–D shows that the post-workshop TTFA profiles (Table 10C) and pre-workshop MLR(T)/FA profiles (Table 10A) both are reasonably consistent with the true profiles (Table 10D). For the unknown profile (sandblast), the normalized residuals (i.e. the ratios of deviations to estimated standard errors) range from −3.2 to +2.0 for TTFA (post-workshop) and from −3.6 to +5.6 for MLR(T)/FA. The pre-workshop TTFA profile estimates are more deviant, with residuals ranging from −9.2 to 6.2. It is noteworthy that TTFA correctly identified the largest number of source classes: 7 out of 9 in the post-workshop solution, and 6 out of 9 in the pre-workshop solution. Each of the other methods identified five classes.

(6) The quality of fit for source impact estimates, as reflected in the normalized residuals, is roughly comparable for all methods (see Table 12). Excluding the road estimates, the maximum absolute normalized residuals range from 4.5 to 16. Methods that inferred their own profiles [MLR(T)/FA and TTFA] show large negative residuals for road estimates (−10.5 and −34, respectively).

There are several possibilities for judging the overall success of a particular source apportionment analysis. For example, a graphical approach is used in connection with the Set II results (Fig. 5). As a numerical alternative, we define a 'figure of merit', F

$$F = \frac{1}{P}\sum_i \text{minimum}\left[\frac{\text{Calc}}{\text{True}}i, \frac{\text{True}}{\text{Calc}}i\right]. \tag{6}$$

where Calc and True are the predicted and actual intensities of the ith source, and the summation is over all P sources contributing to the receptor site. F lies in the range between 0 and 1, and is equal to 1 if and only if there is perfect agreement; a value of 0.5 indicates that the predicted and actual intensities agree within an average factor of 2. There are admittedly some deficiencies in such a definition: as an averaged measure of agreement, F may be greatly influenced by the poor agreement of weak source intensities; F takes no account of the magnitudes of a solution's calculated uncertainties; and F is subject to significant uncertainty due to the relatively few degrees of freedom. Bearing these limitations in mind, for the five solutions shown in Table 12, we find that F ranges from a best value of 0.76 to a worst value of 0.42, with an average of 0.54. From the latter value we can state that the source apportionment procedures were able to predict source impacts (for the sources resolved) within a factor of 2 on average.

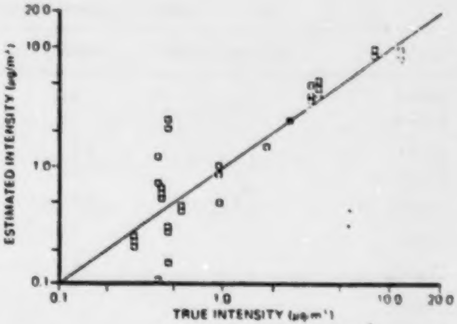

Fig. 5. Estimated vs true Set II source intensities for CMB(EV-1)$_1$, CMB(RR)$_2$ and CMB(EV-2) solutions.

Finally, it is of interest to consider whether the calculated uncertainties for the Set I solutions are reasonable in their magnitudes. Since the uncertainties are intended to represent one standard deviation, the differences between the source intensity estimates and the true values should not exceed ~3 times the stated uncertainties and, in the long run, ~68 % should lie within the stated uncertainties. All sets of results are consistent with this expected proportion (68 %), with the second TTFA solution lying closest, but there are too few degrees of freedom for a very sensitive test. (Because of sampling size error, with just seven comparisons, the proportion could range from 2/7 to 7/7 by chance [90 %, confidence interval].) All solutions have normalized residuals (differences from the truth/stated uncertainties) that are larger than expected; CMB(EV)/FA is best in this regard (residuals of −4.5 to 4.1) and TTFA (pre-workshop) is widest (residuals of −41 to 5.9). It follows that the estimated uncertainties were generally too small.

Set II results

Table 13 contains results for Set II that were submitted both before and after the workshop. The pre- and post-workshop CMB(RR) results differ in how the data were weighted in the fitting process. In the former, the weight for an average elemental concentration took into account both its measurement standard error and its day-to-day variability. In the latter, the weight was based on the quoted concentration (measurement) error only. Also, in the latter case, the criteria used to choose the ridge regression k-value led to selecting $k = 0$. In other words, the best solution corresponded to ordinary weighted least-squares CMB. Thus, the general agreement between the CMB(RR)$_2$ solution and CMB(EV-1)$_2$ is to be expected, since both resulted from averaging the elemental concentrations, and since the difference between ordinary and effective variance weighting is probably not large in this case. Similarly, the close agreement between CMB(EV-1)$_1$ and CMB(EV-2) is to be expected.

Not shown in Table 13 is a TTFA 4-source solution obtained prior to the workshop: auto = 6.2 ± 0.2 μg m^{-3}, incinerator = 3.2 ± 0.1 μg m^{-3}, coal = 5.3 ± 0.3 μg m^{-3}, soil = 19.6 ± 0.3 μg m^{-3}. This approach thus significantly underestimated the number of major sources present in Set II. Little improvement was accomplished in post-workshop attempts with this method. The author who contributed this solution (PKH) assumes that the cause of these difficulties in source resolution was that many of the sources were collocated, such that their variation was governed predominantly by the dispersion model.

The figures of merit calculated from Equation (6) for the five solutions given in Table 13 range from a best value of 0.76 to a worst value of 0.57, with an average of 0.67. On this basis, the post-workshop CMB(EV-2) solution was closest to the true source intensities, but it was only marginally closer than the pre-workshop CMB(EV-1)$_1$ solution. On the basis of the figure of merit values, we conclude that these source apportionment procedures were able to predict source impacts to better than a factor of 2 on the average.

A graphical alternative to the numerical figure of merit is given by Fig. 5, in which the Set II results for CMB(EV-1)$_1$, CMB(RR)$_2$ and CMB(EV-2) are compared with the true values. Because of the use of logarithmic axes, the vertical linear distance of each point from the straight line is proportional to the percentage deviation of the result from the true value. Notice that, because of collinearity among source profiles, the quality of agreement does *not* improve monotonically with increasing intensity of a source.

A striking feature of the results presented in Table 13 is that while there is generally good agreement among intensity estimates for a given source, the calculated uncertainties are widely different, ranging from very small uncertainties for CMB(EV-1)$_1$ to very large for CMB(RR). If we examine the uncertainties listed in Table 13 as we examined those in Table 12, we find that only the CMB(EV-1)$_2$ and CMB(EV-2) uncertainties are consistent with the deviations. The CMB(EV-1)$_1$ estimates and their uncertainties almost never overlap the true values, and the CMB(RR) results always do. (For a true standard deviation, the expected overlap probability is ~68%. However, because of sampling size error, with 13 comparisons, this proportion could range from 5/13 to 11/13 by chance [90% confidence interval].) In light of the expected similarity in the CMB(EV-1)$_2$ and CMB(RR)$_2$ solutions (discussed earlier), the large difference in relative uncertainties between these two is inconsistent. Also, the CMB(RR) uncertainties are so large (RSD > 60%) that eight sources must be considered 'undetected' by that method.

Set III observations

No results for Set III were submitted by any participant, but it was evident (because of the additional log-normal error components) that deconvolution would have been still more difficult than for Set II. The importance of additional physicochemical characterization of source and ambient samples for these difficult but realistic deconvolution situations is evident from the demonstrated effects of collinearity among source profiles. An explicit illustration of this was given by the National Bureau of Standards, where the effect of ^{14}C characterization was investigated. For Set III, without ^{14}C measurements, there were 16 very strong pairwise correlations among the 13 source estimates; the addition of ^{14}C characterization reduced that number to only 1 severe interference (between steel-A and steel-B). (These interference questions were addressed mathematically by examining the variance-covariance matrix resulting from weighted least-squares analysis.)

Carbon-14 characterization was obviously crucial in reducing the interference between wood and automotive sources. Clearly, other correlation-breaking observations, such as additional isotopic analyses and individual particle and crystalline phase analyses, may be the key for resolving a host of sources with adequate precision for control purposes.

UNCERTAINTIES

One topic that deserves further investigation in future intercomparisons and methods development is the matter of uncertainties. It is quite encouraging (and represents real progress) that all of the reported results included approximate standard errors. Some obvious differences in treating different error components, degrees of freedom, and confidence intervals were seen, but an attempt was made to quote all results on an equivalent basis. Limitations in current treatments (especially with respect to systematic error and the treatment of negative estimates) are summarized below.

(1) None of the reported results included a completely adequate statement of uncertainties. Comparisons of deviations from the truth with error estimates (particularly for the results shown in Table 13) confirm that different investigators employed different bases for stating uncertainties.

(2) For methods using the profiles as submitted (some form of CMB): only the effective variance CMBs explicitly included the effect of source profile (σ_m) uncertainties, but these were treated as random rather than as systematic error components.

The use of effective variance itself is an approximate treatment of the total least-squares problem (errors in the source matrix as well as in the observations); a more rigorous approach is given by Golub and Van Loan (1980). Still more approximate, however, were CMBs which explicitly included the observation measurement error (σ_i) only. This was generally by choice, however, and not indicative of an inherent limitation in the method (e.g. ridge regression).

(3) Multivariate and factor analysis solutions quoted uncertainties that were to be taken as no greater than the standard error; no information on bias bounds (if any) was provided.

(4) Ridge regression provided no quantification of bounds for the bias that is introduced by this mathematical method. However, according to the author who contributed this solution (HJW), the absence of significant negatives or unreasonably large positive values precludes the possibility of extreme biases.

(5) Uncertainty about such items as the appropriate number of degrees of freedom prevented the construction of confidence intervals based on Student's *t*.

(6) Consistency tests between the MLR(T)/FA and TTFA source profile estimates and the true profiles were likewise difficult because of insufficient information about uncertainties in the estimated profiles.

(7) For Set I (fewer than 13 sources operating), the questions of uncertainty intervals for the estimated number of sources or alternative possible solutions within such intervals were generally not addressed.

(8) Quality of fit measures would have been desirable for all methods and results. In at least some cases discussed at the workshop, it was apparent that such measures (e.g. chi-square) did not always remain within statistically acceptable bounds.

CONCLUSIONS

Despite the extremely short time available for analysis prior to the workshop, the methods of deconvolution were generally satisfactory, providing consistent estimates for nearly all sources present. In quantitative terms, the best solutions for the two data sets that were analyzed gave source intensity estimates that averaged within ~ 30% of the true values. It was not obvious that any one method was 'best' in all circumstances. Specifically, the multivariate methods (factor analyses and multiple linear regression by tracers) were very successful in dealing with Set I, which contained a relatively small number of sources, one of whose profiles was unknown. On the other hand, the least-squares (CMB) method was clearly the better choice for Set II, which contained a large number of active sources with no unknown source profiles. Using these observations in the source apportionment of a real data set would be difficult because a real data set is likely to contain a large number of sources, at least some of which probably are unknown. That is, a real data set would probably combine the more difficult aspects of Sets I and II. It seems apparent that there is value in applying both CMB and factor analysis approaches to a given data set, and insisting on some degree of consistency between them before regarding the results as valid.

Time was insufficient for full use of meteorological data by the various methods. The two factor analyses that did use meteorological variables had different degrees of success. Given the limited total number of 'samples' available for analysis, no participant attempted to use the meteorological data to stratify the ambient concentration data prior to analysis. This approach might have been useful if the participants had been supplied with the source geography (Figs 1 and 2) and with more observations.

Limitations in the analysis of the data sets related primarily to limitations in power of resolution and to different (or incomplete) approaches to stating uncertainties. The latter limitation has already been discussed. The former is intrinsic in the data (*not* the deconvolution methods) and is expected to be solved by adding correlation-breaking data (e.g. isotopic and microscopic), which tend to carry unique information about different sources. Sampling regimes that permit the application of such techniques as wind trajectory analysis may also be very effective in reducing the effects of profile collinearities and in discovering superpositions (such as 'road' in the current study). The eventual marriage between deterministic (dispersion and source inventory) and stochastic (receptor-deconvolution) models may also lead to greatly improved resolution together with identification-quantification accuracy. Whether either model is sufficiently advanced for an effective linkage is unclear at present; this linkage would seem to be a matter for further research. (See comments below concerning the limitations of the RAM dispersion model.)

Strengths and weaknesses in the simulated data, *per se*, were evident in the generation process and in the results of the intercomparison. The greatest strength was the fact that rigorously controlled models, with known error structures and source concentrations, could be applied. The ranges of concentrations and source types were reasonable (realistic) and 'interesting' because participants could readily quantify some components but scarcely detect others. The unknown number of components, the 'extra' component, the linear combination component, and the collinearity of profiles served very useful purposes, by representing situations that can be expected to occur in real field data. The need for more attention to treating the different error classes and estimating uncertainties was usefully demonstrated by the simulated data, as was the need for additional kinds of physicochemical information (if one hopes to resolve cases as difficult as Set III).

The exercises were not, of course, broad enough. Having more extensive world lists of sources, data sets containing more than a single unknown source, and multiple receptors [illegible] additional wind direction data would be desirable. More observation periods also would be desirable, since the number of degrees of freedom in each data set was barely adequate to properly apply factor analysis (Henry *et al.*, 1984). The greatest limitations in the simulation process were various peculiarities of the RAM dispersion model. The narrowness of the dispersion led to extreme fluctuations in source contributions at the receptor. Also, this narrowness together with the RAM's different treatments of point and area sources and lack of a subroutine for line sources (highways) led to unrealistic (too large, too variable) emissions strengths for several source types.

Based on all of the foregoing considerations (apportionment techniques, simulation method, and data), we conclude that future simulation exercises should consider the following improvements.

(1) Generate replicate observations to provide *direct* measures of imprecision and bias.

(2) Overcome plume narrowness by using an improved dispersion model.

(3) Provide participants with geographical information for both sources and receptors.

(4) Use multiple (~ 3) receptors to evaluate the ability to pinpoint unknown sources and estimate upper limits of emissions from specified sources.

(5) Increase potential resolving power by providing not only hourly wind directions, but also information on direction variability (σ_θ) for each hour.

(6) Treat σ_H as a discrete (feedstock) change, rather than as random with each sampling period.

(7) Change sampling times from 1200–2400 to 0700–1900 to better expose diurnal variations.

(8) Introduce unknown sources to test for false positives and negatives (and for profile estimation) by including ≥ 2 unknown sources and expanding the world list.

(9) Use both fine and coarse fractions.

(10) Include morphological data.

(11) Include sensitivity studies of source intensity detection threshold and source resolution of similar profiles.

(12) Increase the number of sampling periods to meet the statistical requirements of factor analysis.

Acknowledgements—RWG gratefully acknowledges support for part of this work through a Fellowship awarded by the National Research Council.

REFERENCES

Bevington P. R. (1969) *Data Reduction and Error Analysis for the Physical Sciences.* McGraw-Hill, New York.

Bowne N. E., Cher M. and Liu M. (1981) EPRI Plume Model Validation Project: preliminary evaluation of model performance at a Plains site. In *Proc. 74th Annual Meeting and Exposition of the Air Pollution Control Association*, Philadelphia, Pennsylvania, pp. 1–16 (Session 20, Paper 1).

Currie L. A. (1977) Detection and quantitation in X-ray fluorescence spectrometry. In *X-Ray Fluorescence Analysis of Environmental Samples* (edited by Dzubay), pp. 289–306. Ann Arbor, Ann Arbor, Michigan.

Dzubay T. G., Stevens R. K., Courtney W. J. and Drane E. A. (1981) Chemical element balance analysis of Denver aerosol. In *Electron Microscopy and X-ray Applications to Environmental and Occupational Health Analysis*, Vol. 2 (edited by Russell), pp. 23–42. Ann Arbor, Ann Arbor, Michigan.

Golub G. H. and Van Loan C. F. (1980) An analysis of the total least squares problem. *SIAM J. Numer. Anal.* 17, 883–893.

Henry R. C., Lewis C. W., Hopke P. K. and Williamson H. J. (1984) Review of receptor model fundamentals. *Atmospheric Environment* 18, 1507–1515.

Novak J. H. and Turner D. B. (1976) An efficient Gaussian-plume multiple-source air quality algorithm. *J. Air Pollut. Control Ass.* 26, 570–575.

Parr R. M., Houtermans H. and Schaerf K. (1979) The IAEA intercomparison of methods for processing Ge(Li) gamma-ray spectra. In *Computers in Activation Analysis and Gamma-Ray Spectroscopy*, pp. 544–562. CONF-780421, National Technical Information Service, Springfield, Virginia.

Roscoe B. A. and Hopke P. K. (1981a) Comparison of weighted and unweighted target transformation rotations in factor analysis. *Comput. Chem.* 5, 1–7.

Roscoe B. A. and Hopke P. K. (1981b) Error estimates for factor loadings and scores obtained with target transformation factor analysis. *Analytica chimica Acta* 132, 89–97.

Stevens R. K. and Pace T. G. (1984) Overview of the Mathematical and Empirical Receptor Models Workshop (Quail Roost II). *Atmospheric Environment* 18, 1499–1506.

Turner D. B. and Novak J. H. (1978) *User's Guide for RAM.* EPA-600/8-78-016a, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Atmospheric Environment Vol. 18, No. 8, pp. 1539–1553, 1984
Printed in Great Britain

[illegible]
1984 Pergamon Press Ltd.

CHEMICAL AND PHYSICAL ANALYSES OF HOUSTON AEROSOL FOR INTERLABORATORY COMPARISON OF SOURCE APPORTIONMENT PROCEDURES*

DAVID L. JOHNSON†
Department of Chemistry, State University of New York College of Environmental Science and Forestry, Syracuse, NY 13210, U.S.A.

BRIANT L. DAVIS
South Dakota School of Mines and Technology, Rapid City, SD 57701, U.S.A.

THOMAS G. DZUBAY, HASHIMA HASAN
U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, U.S.A.

E. RUSSELL CRUTCHER
Microlab Northwest, Redmond, WA 98052, U.S.A.

WILLIAM J. COURTNEY
Northrop Services, Inc. Research Triangle Park, NC 27709, U.S.A.

JOSEPH M. JAKLEVIC, A. C. THOMPSON
Lawrence Berkeley Laboratory, Berkeley, CA 94720, U.S.A.

and

PHILIP K. HOPKE
Institute for Environmental Studies, University of Illinois at Urbana-Champaign, Urbana, IL 61801, U.S.A.

(*First received* 9 *June* 1983 *and in final form* 19 *January* 1984)

Abstract—Eighteen consecutive 12-h ambient aerosol samples were obtained in Houston, TX, as part of a U.S. Environmental Protection Agency visibility study in that city. The samples were collected on Teflon, Nuclepore and quartz filters by means of dichotomous samplers and were analyzed for mass, elemental and ionic species concentrations by β-ray attenuation, X-ray fluorescence, ion chromatography, colorimetry, instrumental neutron activation analysis and pyrolysis. The specimens and analytical data were distributed to participants of the Mathematical and Empirical Receptor Models Workshop (Quail Roost II), to provide a basis for comparison of source apportionment procedures. Certain workshop participants undertook further analyses of the specimens by X-ray powder diffraction, scanning electron microscopy with automated image analysis and X-ray energy spectroscopy, and analytical light microscopy. This report describes all of these analytical procedures. Also, through prototype comparison mechanisms, the elemental, compound and individual particle data are reduced to a common basis (elemental composition), permitting intercomparisons of results from different methods. In general, the intercomparisons suggest that each analytical technique has certain advantages and limitations for receptor modeling and that the various techniques are best viewed as complimentary. Combinations of analytical approaches can lead to improved source resolution in the apportionment calculations.

INTRODUCTION

The application of receptor models to atmospheric particle source apportionment is based on the physicochemical characterization of ambient aerosol. The use of such models constitutes an attempt to project backwards in time and space to the sources of this collected material. In addressing the impact of particles on the consumers of air resources, both quantitative and qualitative data are required – we wish to know, of course, 'how much' of 'what' comes from 'where.' Several chemical analytical methods may be employed to quantitatively describe atmospheric aerosols. Receptor models mathematically treat the obtained descriptions to quantify defined sources. Thus, sampling of ambient aerosols and source materials, characterization of the samples, and data treatment are equally critical to receptor modeling.

---

* A report from the Mathematical and Empirical Receptor Models Workshop (Quail Roost II).

† To whom requests for reprints of this article should be addressed. Requests for reprints of the entire series of six articles (including this one) documenting the Quail Roost II workshop should be directed to Robert K. Stevens, MD-47, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, U.S.A.

This report describes the collection and characterization of Houston, TX, aerosol as an example of 'state-of-the-art' ambient aerosol sampling and analysis. The resulting data set provided one of the bases for comparison of source apportionment procedures in the Mathematical and Empirical Receptor Models Workshop (Quail Roost II) (Stevens and Pace, 1984).

Existing source apportionment models rely heavily on the measurement of elemental concentrations of ambient atmospheric particles. These data are generally obtained using reliable and well-understood analytical methods such as X-ray fluorescence (XRF) and instrumental neutron activation analysis (INAA). Such methods are cost-effective and have proven successful in several previous receptor model applications. But the number of emission sources that can be resolved on the basis of bulk elemental composition is limited by the multicollinearities that occur in the source signatures if many source profiles are included in the analysis. Greater resolution of sources may be achieved if additional analytical methods, such as X-ray powder diffraction (XRD) for quantification of crystalline components, and microscopy to characterize individual particles, are used. We shall discuss all of these techniques and evaluate their utility for receptor modeling.

EXPERIMENTAL

During September 1980, a detailed set of measurements was made to determine relationships among the optical, physical and chemical properties of aerosol in Houston (Dzubay *et al.*, 1982). In that study, aerosol samples were obtained by four dichotomous samplers operated simultaneously to collect particles in the fine (0.1–2.5 μm) and coarse (2.5–15 μm) aerodynamic diameter size ranges. Table 1 summarizes the filter media, species measured and analytical procedures. To enhance the data set for use in receptor models, the analyses included INAA, XRD, analytical light microscopy (ALM), and scanning electron microscopy with automated image analysis and X-ray energy spectroscopy (SAX). Table 2 shows the schedule for all analyses and the methods of sample transport among laboratories. The sampling site and sampling and analytical methods are described briefly below.

Site

The measurement site was on the University of Houston campus. Three automated dichotomous samplers were located in a shelter at ground level at the north end of the campus, and two additional aerosol samplers and two vans containing instruments were located within 20 m of the shelter. Grassy areas and campus buildings were to the east, south and west of the shelter.

Sampling methods

Four dichotomous samplers were used to collect particles in the 0.1- to 2.5-μm (fine) and 2.5- to 15-μm (coarse) ranges of aerodynamic diameter. Three were automated dichotomous samplers (Beckman Instruments, Inc.). Their filters were mounted in 5- × 5-cm square frames and held in trays in groups of 36. The Beckman samplers were set to change filters [illegible] and [illegible] CDT, and they were programmed to make

Table 1. Summary of sampling devices and analytical procedures used to characterize Houston aerosol

Sampler type	Manufacturer (model)	Height of inlet (m)	Filter media	Species measured	Analytical procedure(s)*
Dichotomous	Sierra (244)	2	Teflon†	NO_3^-, SO_4^{2-}	Ion chromatography
				H^+	Gran titration
				NH_4^+	Colorimetry
Dichotomous	Beckman	4	Teflon†	Mass	β-gauge
				Elemental composition	XRF, INAA
				Crystalline components	XRD
Dichotomous	Beckman	4	Nuclepore‡	Elemental composition	XRF
				Individual particles	SAX, ALM
Dichotomous	Beckman	4	Quartz§	Carbon	Combustion
Hi vol (< 4 μm)	Prototype	2	Quartz§	Carbon	Combustion
				^{14}Carbon	β-counter

* Definitions of acronyms: XRF, X-ray fluorescence; INAA, instrumental neutron activation analysis; XRD, X-ray diffraction; SAX, scanning electron microscopy with automated image analysis and X-ray energy spectroscopy; ALM, analytical light microscopy.

† 1 μm.

‡ 0.3 μm.

§ QAST 2500.

Table 2. History of sample handling and analyses

Filter medium	Method*	Analysis Location†	Date(s)	Transport method to analysis site
Teflon	β-gauge	RTP, NC	24 Sept. 1980	Hand carried
	XRF	RTP, NC	6 Oct. 1980	
	Light transmission	RTP, NC	8 Oct. 1980	
	Photoacoustic	Raleigh, NC	10 Oct.–9 Dec. 1980	Hand carried
	β-gauge	RTP, NC	1 Jan. 1981	Hand carried
	Storage in desiccator	RTP, NC	1–22 Jan. 1981	
	β-gauge	RTP, NC	22 Jan. 1981	
	Storage in desiccator	RTP, NC	23 Jan.–20 Feb. 1981	
	β-gauge	RTP, NC	23 Feb. 1981	
	Photoacoustic	Raleigh, NC	24 Feb.–4 Aug. 1981	Hand carried‡
	XRF	Berkeley, CA	1 Oct.–2 Oct. 1981	U.S. mail§
	XRD	Berkeley, CA	20 Sept.–2 Oct. 1981	
	XRF	RTP, NC	28 Oct. 1981	U.S. mail§
	INAA	Urbana, IL	1–25 Jan. 1982	U.S. mail‖
Nuclepore	β-gauge	RTP, NC	24 Sept. 1980	Hand carried
	XRF	RTP, NC	7 Oct. 1980	
	(No analysis)	Raleigh, NC	10 Oct. 1980–17 Sept. 1981	Hand carried
	XRF	RTP, NC	17 Sept. 1981	Hand carried
	SAX	Syracuse, NY	1981 Oct.–1982 Jan.	U.S. mail§
	ALM	Redmond, WA	8 Dec. 1981–1982 Jan.	United parcel service§

* Definitions of acronyms: XRF, X-ray fluorescence; XRD, X-ray diffraction; INAA, instrumental neutron activation analysis; SAX, scanning electron microscopy with automated image analysis and X-ray energy spectroscopy; ALM, analytical light microscopy.

† RTP = Research Triangle Park.

‡ Only fine-fraction filters were transported.

§ Filters were shipped in a double-padded box.

‖ Filters were shipped in a triple-padded box.

additional changes whenever clogging caused a 5% decrease in flow rate. Never more than one Teflon or quartz filter was used for a 12-h sampling period, but clogging often resulted in the use of two Nuclepore filters. Samples for ionic analysis were collected in a manually operated dichotomous sampler on 37-mm diameter, 1-μm pore size Teflon membrane filters bonded to polyolefin rings (Ghia Corp.). Filters were changed every 12 h at approximately 0700 and 1900 CDT. Immediately after sampling, the filters were placed, individually, in 30-ml polypropylene bottles that contained 25 ml of extraction solution (10^{-5} N ammonium chloride and 5×10^{-5} N perchloric acid). The inlet flow rate of each dichotomous sampler was set at 16.7 l min^{-1}, and the fraction of inlet air flow passing through the coarse particle filter was set at 0.1 using a dry test meter.

A prototype sampler was used to collect fine particles on 20- × 25-cm Pallflex 2500 QAST quartz filters for subsequent ^{14}C analysis. The inlet consisted of six coaxial cylinders that provided a tortuous path to remove coarse particles by impaction. The upper 50% cut point was an aerodynamic diameter of 4 ± 1 μm. A brushless pump motor and a pneumatic servomechanism were used to maintain the flow rate at 980 l min^{-1}. Sampled filters and several blanks were folded, wrapped individually in aluminum foil, and placed in paper envelopes.

Analytical methods

The techniques used to determine mass, bulk elemental composition, and bulk ionic composition have been reported previously (Dzubay *et al.*, 1982), and only a brief summary is given below. The subsequent analyses by INAA, XRD, SAX and ALM are described in greater detail.

β-ray attenuation. Mass was measured by β-ray attenuation using an instrument and procedures described by Jaklevic *et al.* (1981). The precision was ± 2.7 μg m^{-3} (one standard deviation) for 12-h sampling periods and 6.5-cm^2 deposits. From a comparison of β-gauge and gravimetric measurements (Courtney *et al.*, 1982), we infer that any bias in the mass measurements by β-gauge was < 5%. The relative humidity (r.h.) during the initial and final mass determinations was 53 and 57 ± 3%, respectively.

X-ray fluorescence. All samples collected on Teflon or Nuclepore filters were subjected to XRF analysis for bulk elemental composition (Dzubay, 1977a, b). Concentrations of elements with atomic numbers of > 12 were determined using an energy dispersive XRF spectrometer (Jaklevic *et al.*, 1976). The X-ray spectra were analyzed using a least-squares fitting procedure (Drane *et al.*, 1980). Calibration for sulfur was based on pure ammonium sulfate aerosol deposits, which were analyzed gravimetrically and verified by wet chemical analysis (Shaw *et al.*, 1980). For all other elements, pairs of gravimetrically analyzed vapor-deposited metal films (Micromatter Co.) were used as standards. Corrections for attenuation of soft X-rays in coarse particles were made using factors described previously (Dzubay and Nelson, 1975).

Ion chromatography and colorimetry. The samples stored in extraction solution were prepared for ionic analysis by ultrasonic agitation for 1 h. Sulfate, nitrate, and sodium were determined in 2-ml aliquots by ion chromatography (IC) using anion and cation columns. Ammonium was determined by an indophenol blue method using a modified Technicon autoanalyzer.

Instrumental neutron activation analysis. For INAA, the Teflon filters were cut from their support frames and packaged in nitric-acid-cleaned polyethylene bags. Standards were prepared by spotting onto filter paper known volumes of multielemental solutions. The solution standards were checked by analyzing samples of NBS SRM 1633, Trace Elements in Fly Ash. The samples were first analyzed for very-short-lived activity by irradiating them one at a time in the

Pneumatic Tube Facility of the Illinois Advanced TRIGA at a flux of 1×10^{13} neutrons $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ with a nickel foil flux monitor. After a 2-min decay time, the samples were counted for 200 s at a fixed geometry relative to a lithium-drifted germanium detector. The programs described by Maney *et al.* (1977) were used to convert the spectral data to elemental concentrations of Al, Ti, V, Mg and Cu.

Ten samples and two sets each of standards for short-lived isotopes (Mn, In, Dy, Ba and Sr) and intermediate-lived isotopes (Na, K, Ga, Sb, As, Eu, La and Au) were irradiated in the Rotating Facility of the Illinois reactor for 1 h at a flux of 2×10^{12} neutrons $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ and a cadmium ratio of ~ 3. The samples and standards were returned to the laboratory, transferred to clean, unirradiated polyethylene bags, and then sequentially counted at a fixed geometry using a sample changer described by Roscoe (1981) for 900 s between 1 to 4 h after irradiation. They were recounted for 1 h each beginning 24 h after the end of the irradiation. The data were collected and stored in a computer-based multichannel analyzer, where programs converted the gamma ray spectral data into elemental concentrations.

Pyrolysis. Volatile and total carbon were determined by pyrolysis (Stevens *et al.*, 1982) using a Model DC-50 carbon analyzer (Envirotech Corp.). Volatile carbon was determined on 6.7-mm dia. portions of quartz filters by pyrolysis in a pure helium atmosphere in a quartz tube at 630 °C. Total carbon was determined similarly, except that the samples were imbedded in manganese dioxide granules, and the pyrolysis temperature was 850 °C.

X-ray powder diffraction. The strength of XRD is its ability to quantify the crystalline composition of ambient aerosol samples. Provenance analysis (Davis *et al.*, 1981; Davis *et al.*, in press) based upon XRD results and appropriate area source information has been used directly in source apportionment schemes. For the present work, a complete set of aerosol samples collected on Teflon filters was analyzed at Lawrence Berkeley Laboratory (LBL). The instrument was an automated system using an 800-W Cu X-ray tube and a position-sensitive detector (Thompson *et al.*, 1982). This system increases the data acquisition rate by a factor of ~ 50 over a conventional single-slit system. Each filter was scanned at LBL from 5.5° 2Θ (Cu K_α radiation) to 14° 2Θ at a constant 5-mm × 10-mm irradiation area and from 14 to 50° at a 15-mm × 10-mm irradiation area. The scan time was $750\ \mathrm{s}\ {}^{\circ}2\Theta^{-1}$.

The resulting XRD spectra were analyzed at LBL using spectral stripping routines in an on-line computer. After a normalized background was subtracted, peak positions and areas were determined. Then a library of 77 XRD patterns for common species in aerosol was used in a search-match procedure to identify crystalline species in each sample. Absolute concentrations (independent of the presence of noncrystalline species) of the identified species were determined using calibration constants deduced from spectral standards acquired and analyzed in a similar fashion. The standards were prepared by using ground powders of pure minerals and collecting samples onto Teflon filters with a dichotomous sampler. The amount of standard material on the Teflon sample was measured with a β-gauge and verified, where possible, by XRF. The calibration factors had an accuracy of ~ 20%. Component detectability of the LBL instrument was ~ 5 $\mu\mathrm{g\,cm}^{-2}$.

One of the authors (BLD of the South Dakota School of Mines and Technology [SDSM]) obtained plots of the XRD patterns produced on the LBL diffractometer and performed an independent analysis based on those spectra. The analysis consisted of first identifying crystalline species on the basis of peak heights and intensities. Then the fractional abundances of the crystalline species were determined by a reference-intensity method (Davis, 1978, 1980, 1981a). The analysis included correction for transparency and matrix effects. Analytical uncertainties for these SDSM crystalline determinations ranged from 5 to 10% for major components to > 100% for minor components. A more complete description of the error analysis and values for the Houston data set is given by Davis (1981b, 1982).

Soil and road dust samples obtained in Houston were analyzed by XRD at SDSM. Sample preparation consisted of suspending the soil-dust using an aerosol generator and collecting the suspended particles on a filter. X-ray transmission (XRT) measurements were made on blank glass-fiber filters, and similar measurements were completed on the loaded filters in order to obtain an estimate of amorphous material (assumed to be carbon-rich, with a mass absorption coefficient of $8.0 \pm 2\,\mathrm{cm}^2\,\mathrm{g}^{-1}$). The diffraction spectra for these samples were obtained at a scan rate of $120\ \mathrm{s}\ {}^{\circ}2\Theta^{-1}$ with a minimum irradiation beam length of 12 mm. Quantitative amounts of crystalline species were determined by the reference-intensity method.

In the SDSM and LBL analyses, the weight of each component was reported as the oxide or element by use of stoichiometric coefficients and average component compositions given in compilations of Deer *et al.* (1962). Calculations include the presence of minor substitution elements that may not ordinarily appear in a mineral's chemical formula. The final list of oxides or elements, therefore, represents the 'most probable' composition expected on the basis of the average mineral extracted from the Deer *et al.* tables.

Scanning electron microscopy with automated image analysis and X-ray energy spectroscopy. SAX utilizes the fact that all particles are unique. Individual particle analysis allows quantitative characterization of this aerosol heterogeneity. Computer-interfaced scanning electron microscopic methods for particle characterization permit the categorization of thousands of particle 'object vectors' h^{-1} (Lee and Kelly, 1980; Hanna *et al.*, 1980; Johnson *et al.*, 1981; Johnson, 1983). Thus, the obtained information on particle heterogeneity is amenable to the several types of statistical manipulation used in traditional forms of sample analysis (factor analysis, linear regression, etc.).

SAX software systems interfaced with scanning electron microscopic and X-ray energy spectroscopic hardware create descriptions for each particle or feature that is analyzed. These descriptions are the loci of particles on the hyper-ellipsoid comprised of several morphological measures, such as width, length, average diameter, area, etc., and individual particle chemical element measures based on X-ray acquisition. Classification schemes can categorize these 'object vectors' and place them into particle type classes having some geochemical reality. Categorization schemes are built upon both *a priori* knowledge and empirical information. SAX results from a signature histogram of population percent or weight percent versus particle species. Source and ambient aerosol signatures are produced in the same fashion; a process directly analogous to the production of signatures by bulk chemical elemental composition, but with the observation variables for a sample partitioned as physicochemical particle types instead of as elemental concentrations.

Samples for SAX were prepared from the Nuclepore membranes by cutting a 1-cm^2 section of each filter halfway between the edge and the center. The membrane subsections were cemented to graphite stubs with colloidal graphite in isopropanol. Before analysis in the scanning electron microscope, the specimens were coated with 20–30 nm of carbon in a high-vacuum evaporator. Once in the microscope, random fields of view were selected for analysis. At least 500 particles were characterized for each sample; for many samples, 1000 particles were analyzed. Particles were characterized at several magnifications, allowing occasional stray large particles (optical diameter > 50 μm) to be analyzed.

Analytical light microscopy. For ALM, the Nuclepore filters were examined in a laminar-flow clean bench to evaluate particle distribution. A photograph was taken to document this distribution, and a sample area (~ 0.3 cm^2) on the filter selected. Particles were lifted from the selected area with a fine tungsten needle wetted with Aroclor 5442, and

Reproduced from
best available copy.

another photograph was taken to document collection bias. The particles were then transferred to a heated slide (70 C) and mounted in Aroclor 5442 (refractive index 1.66). The coverslip was added, and the sample was left for 1 h to equilibrate.

A modified qualitative analysis was performed on the mounted sample to establish a set of optical categories. The optical criteria used for this analysis were part of a master list used to generate the criteria for any analysis. Then, random fields from the mounted sample were selected for analysis until a total particle count (for all fields) of > 1000 was reached. The analytical 'field' consisted of all separate particles within the boundaries of a modified Porton reticule. Particles were considered separate unless bordered by a matrix that contained other individually enclosed particles. Particles were sized by comparison to the set of circular areas on the Porton reticule. The projected area of the particle was used for the sizing, and the resultant data were converted into a volume by category (spherical assumption). A density was assigned to each category and a mass by category calculated. These data were then used to derive the percentage by mass.

RESULTS AND COMPARISONS OF ANALYTICAL TECHNIQUES

The average aerosol elemental composition data sent to participants of Quail Roost II (Table 3) are similar to those measured earlier in Houston and other locations (Stevens *et al.*, 1978, 1979, 1980). The uncertainties reported represent one standard deviation of the 24 analyses for the sampling period. The carbon data represent total carbon. (Volatile carbon accounted for 80% of the total carbon in the fine fraction.) Elemental carbon was not detected in the coarse fraction; the detection limit was 1.4 $\mu g\ m^{-3}$. Analysis of ^{14}C in seven samples by Currie of the National Bureau of Standards indicated that contemporary carbon in particles of < 4 μm ranged from 18 to 62% of total carbon and had a mean value of 40 ± 15%. The average coarse-fraction aerosol concentration for the 18 Teflon filters used for the Quail Roost II study was 25 $\mu g\ m^{-3}$; individual values ranged from 9.4 to 45.3 $\mu g\ m^{-3}$. The resultant mass per unit area on the membranes was less than optimum for XRD analysis, but, on average, the XRD results showed the coarse fraction to be 25–40% calcite ($CaCO_3$), 10–30% quartz (SiO_2), 5–10% mascagnite ($[NH_4]_2SO_4$), 1–5% dolomite/ankerite ($[Mg, Ca]CO_3/CaCO_3 \cdot [Mg, Fe, Mn]CO_3$), and 1–2% halite (NaCl). Minor amounts of magnetite (Fe_3O_4), muscovite ($KAl_2[AlSi_3]O_{10}[OH]_2$), mullite ($Al_6Si_2O_{13}$), microcline ($KAlSi_3O_8$), thenardite ($Na_2SO_4$) and pyrite ($FeS_2$) were also observed.

The total mass accounting was quite good. The difference between the measured total mass and the sum of the chemical components (Table 3) was 22 ± 10% and 2 ± 20% of the fine and coarse fractions, respectively. This difference is presumed to consist primarily of water associated with the salts and minerals and of hydrogen and oxygen associated with the carbon. The tenacity with which any such water is bound to the fine particle samples was tested by storing them over Drierite (anhydrous calcium sulfate) in a desiccator for 7 weeks and then reanalyzing the mass in a 45% r.h. environment. The average mass decrease was small (2 ± 3%). This insignificant decrease suggests that any water residing on the particles must be bound rather tightly. Moreover, liquid water is not expected to exist at 50% r.h. in the fine fraction of Houston's aerosol because the ionic species consisted primarily of $(NH_4)_2SO_4$, which requires a r.h. of > 80% to be deliquescent. The small amounts of sulfuric acid in the samples were probably neutralized by ambient ammonia in the laboratory before the mass was measured.

Microscopic techniques were applied to describe the samples collected in terms of the individual particle types making up the mass on a filter. Both ALM and SAX were employed, though only one sample was characterized by the former technique. Table 4 shows the ALM results f[illegible] Nuclepore filter 124-502C and the

Table 3. Average composition ($ng\ m^{-3}$) of Houston aerosol in fine and coarse fractions for 0700 10 September to 0700 19 September, 1980*

Species	Fine	Coarse
Mass	38600 ± 300	24800 ± 200
Al	123 ± 60	1093 ± 366
Si	210 ± 32	2990 ± 965
P	28 ± 12	61 ± 34
S	4834 ± 279	397 ± 430
Cl	32 ± 6	366 ± 111
K	119 ± 6	170 ± 20
Ca	155 ± 8	2775 ± 144
Ti	−8 ± 9	36 ± 12
V	−2 ± 5	−5 ± 6
Cr	4 ± 3	7 ± 3
Mn	14 ± 2	21 ± 3
Fe	162 ± 9	604 ± 34
Co	1 ± 2	0 ± 2
Ni	4 ± 1	4 ± 1
Cu	28 ± 3	18 ± 3
Zn	84 ± 5	58 ± 4
Ga	0 ± 1	0 ± 1
Ge	1 ± 1	1 ± 1
As	2 ± 3	0 ± 2
Se	1 ± 1	0 ± 0
Br	55 ± 3	36 ± 3
Rb	0 ± 1	1 ± 1
Sr	1 ± 1	8 ± 1
Cd	0 ± 2	0 ± 2
Sn	2 ± 3	−1 ± 3
Sb	6 ± 3	−2 ± 3
Ba	48 ± 13	91 ± 16
W	0 ± 2	2 ± 2
Hg	−1 ± 2	0 ± 2
Pb	465 ± 23	124 ± 10
SO_4^{2-}	14610 ± 1207	907 ± 165
NO_3^-	593 ± 617	1629 ± 235
NH_4^+	4180 ± 379	< 190
Na	317 ± 254	278 ± 238
C	7100 ± 467	3100 ± 36

* The tabulated results are based upon final β-gauge analysis in September 1980, ionic analyses in September and October 1980, carbon analyses in October and November 1980, and XRF analyses in October 1981. Uncertainties presented are a combination of random and systematic errors, and are roughly equivalent to a 95% confidence level.

Table 4. Results of ALM and SAX analyses of Houston Filter 124-502C for 0700 10 September 1980*

	ALM		SAX	
Class/category	% by class	% by category	% by class	% by category
Minerals	49.5		68	9
Quartz		11.4		21
Calcite		9.0		21
Gypsum		0.5		2
Mullite		—		—
Clay		27.6		36
Other		1.0		—
Combustion	10.4		8	
Fused fly ash		5.5		5
Unfused fly ash		—		—
Soot		4.4		3†
Carbon spheres		—		—
Charred wood		0.4		—
Black irregular		0.1		—
Biologicals	3.3		1	
Plant parts		0.5		—
Pollen and spores		2.6		—
Starch		Trace		—
Other		0.2		1
Industrial and transportation	36.8		23	
Industrial hematite		0.1		—
Rust		1.2		—
Iron scale		3.6		—
Magnetite spheres		3.0		—
Slag		6.4		—
Glass		3.7		—
Hydrated cement		1.2		—
$CaO–CaCO_3$		5.4		11
Secondary industrial		7.8		6
Tire wear		4.1		—
Other		0.3		6

* Results are percent by mass and mass loading by category.

† Carbon not determined by SAX; this is the carbon result obtained by Northrop Services, Inc.

SAX results for the same filter. Clearly, the microscopic analyses provide considerable resolution of this aerosol into components which, in some cases, can be apportioned directly. The SAX results were reported in 25 (slightly different) categories, but are here expressed in terms of the ALM categories. (Agreement between ALM and SAX practitioners on a particle classification scheme would help to reduce such differences.) Results are presented as a mass loading on the filter. It should be noted, however, than an *identical* suite of particles was not compared. Rather, the results in Table 4 represent subsamplings of a common filter.

Comparisons of results for Teflon filters

Several analytical method comparisons were made on the set of Teflon membrane filters shared by the Quail Roost II participants. Sulfur was of considerable concern, as it dominated the mass collected in the fine fraction. However, XRF results for sulfur collected in the fine fraction on one set of Teflon filters agreed very well with IC results for sulfate on a second set of filters. For 16 sample comparisons, a 1 : 1 relation (correlation coefficient of 0.99) was observed.

The samples collected on Teflon filters were analyzed for elemental species by Northrop Services, Inc. (NSI) in October 1980 and again in October 1981. Of the 30 elements measured, the replicate mean values agreed within measurement accuracy for all but phosphorus, chlorine, copper, zinc and bromine in the fine fraction, and for all but copper in the coarse fraction. For phosphorous, chlorine, copper and zinc, this suggests that the reported uncertainty estimates were too small. For bromine, in the fine fraction, there was a 21% decrease between October 1980 and October 1981. Such a decrease may have been due to evaporative loss. In addition, specimen excitation in the XRD procedure, where there is a high X-ray exposure, may lead to volatile element losses (O'Connor *et al.*, 1977).

The alkaline nature of coarse particles and their lower surface-to-volume ratio may account for the fact that relatively little bromine loss occurred in the coarse fraction.

For the replicate analyses 1 year apart, the correlation coefficients exceeded 0.9 for 10 elements in the fine fraction and for 12 elements in the coarse fraction. Although the mean values of arsenic and antimony in the fine fraction and tungsten in the coarse fraction were below the detection limits, their correlation coefficients were quite high due to a single high concentration of each of these elements. For the replicate measurements, the mean values agreed within 10% for 7 elements in the fine fraction and for 12 elements in the coarse fraction.

If sample handling caused any particle loss in the intervening period, the effect would have been most pronounced for the coarse fraction. For elements in the coarse fraction that were measured to an accuracy of better than 10%, the losses were as follows: 7% for calcium, 8% for iron, 5% for bromine and 2% for lead. In summary, we conclude that average losses caused by cross-country shipment and handling during a 1-y period were less than 10% for particles on Teflon filters. Note, however, that the losses mentioned relate to bulk chemistry—an average for the whole filter sample. No pre- and post-mailing XRD or SAX analyses were possible with this particular sample set. Therefore, judging the 'fractionation' of compound classes or particle types caused by sample handling and mailing was not possible.

A cross-laboratory comparison of XRF results for the Teflon filters was carried out in the fall of 1981. Table 5 shows the results of the LBL and the NSI

Table 5. Comparison of Teflon filter XRF results obtained by LBL in October 1981 (Y) with those obtained by NSI in October 1981 (X)*

	$\bar{X}$ (ng m^{-3})	$\bar{Y}$ (ng m^{-3})	$\bar{Y}/\bar{X}$	A (ng m^{-3})	*B*	*R*	*N*
				Fine fraction			
Al	115 ± 60	339 ± 148	—	—	—	—	5
Si	210 ± 32	315 ± 80	1.50	20 (86)	1.41 (0.38)	0.89	18
S	4834 ± 279	6092 ± 400	1.26	−62 (113)	1.27 (0.02)	1.00	18
Cl	32 ± 6	81 ± 15	2.52	90 (66)	−0.27 (2.00)	−0.07	18
K	119 ± 6	122 ± 14	1.03	−13 (9)	1.14 (0.07)	0.99	18
Ca	155 ± 8	147 ± 16	0.95	−31 (12)	1.15 (0.07)	0.99	18
Ti	−11 ± 9	30 ± 18	—	—	—	—	8
V	−1 ± 5	19 ± 12	—	—	—	—	6
Cr	6 ± 3	22 ± 10	3.45	19 (7)	0.54 (0.82)	0.53	8
Mn	19 ± 2	24 ± 6	1.27	7 (3)	0.90 (0.12)	0.98	12
Fe	162 ± 9	158 ± 11	0.97	−33 (7)	1.17 (0.03)	1.00	18
Ni	5 ± 1	9 ± 3	1.85	4 (2)	1.16 (0.27)	0.97	8
Cu	28 ± 3	14 ± 3	0.48	−5 (13)	0.67 (0.46)	0.62	17
Zn	84 ± 5	81 ± 5	0.97	−12 (5)	1.11 (0.04)	1.00	18
Br	55 ± 3	66 ± 4	1.21	−6 (6)	1.32 (0.11)	0.99	18
Sr	1 ± 1	5 ± 2	—	—	—	—	18
Ba	105 ± 13	81 ± 21	—	—	—	—	1
Pb	465 ± 23	517 ± 27	1.11	4 (15)	1.10 (0.03)	1.00	18
				Coarse fraction			
Al	1201 ± 366	896 ± 415	0.75	238 (294)	0.55 (0.24)	0.81	15
Si	2990 ± 965	3114 ± —	1.04	−89 (113)	1.07 (0.04)	1.00	18
P	86 ± 34	115 ± 54	1.34	80 (45)	0.40 (0.43)	0.70	7
S	397 ± 430	538 ± 207	1.36	−85 (78)	1.57 (0.19)	0.98	18
Cl	366 ± 111	556 ± 172	1.52	71 (42)	1.32 (0.09)	0.99	18
K	170 ± 20	217 ± 47	1.28	−14 (41)	1.36 (0.23)	0.95	18
Ca	2775 ± 144	3509 ± 662	1.26	76 (66)	1.24 (0.02)	1.00	18
Ti	39 ± 12	69 ± 18	1.77	35 (15)	0.87 (0.33)	0.82	17
V	−7 ± 6	16 ± 11	—	—	—	—	6
Cr	9 ± 3	26 ± 7	2.99	13 (10)	1.47 (1.05)	0.71	11
Mn	22 ± 3	32 ± 6	1.44	5 (5)	1.19 (0.22)	0.95	17
Fe	604 ± 34	706 ± 37	1.17	−39 (32)	1.23 (0.05)	1.00	18
Ni	5 ± 1	9 ± 3	2.00	4 (2)	1.12 (0.37)	0.90	12
Cu	20 ± 3	13 ± 3	0.64	3 (5)	0.50 (0.25)	0.78	14
Zn	58 ± 4	63 ± 4	1.08	−2 (4)	1.12 (0.05)	1.00	18
Br	36 ± 3	38 ± 2	1.07	0 (3)	1.06 (0.09)	0.99	18
Sr	8 ± 1	13 ± 2	1.59	4 (1)	1.14 (0.15)	0.97	18
Ba	124 ± 16	128 ± 20	1.04	−5 (23)	1.08 (0.12)	0.99	10
Pb	124 ± 10	133 ± 8	1.08	7 (8)	1.02 (0.06)	0.99	18

* $\bar{Y}$, $\bar{X}$ and their combined random and systematic uncertainties are average values for the N samples compared. *R* is the correlation coefficient, and *A* and *B* are determined by linear regression for the equation $Y = A + BX$. Results are shown whenever $N > 5$, $X > 1.5$ ng m^{-3} and $Y > 1.5$ ng m^{-3}.

analyses (average for the 18 filter pairs) for the coarse and fine fractions. An examination of these results shows good agreement and precision in the case of some elements as well as an analytical bias between the LBL and NSI results. The precision is illustrated by the fact that the correlation coefficients exceed 0.9 for 9 elements in the fine fraction and for 13 elements in the coarse fraction. Also, the uncertainty intervals for slope and intercept are fairly small for many elements. However, although the error estimates for individual filters are large enough to span the differences between the two results, there is a significant bias between the two laboratories for some elements. For major elements such as sulfur, potassium, calcium, iron bromine and lead, the LBL results are from 3 to 28% higher than the NSI results. Additional research will be needed to resolve the larger discrepancies, which are mainly for elements of low atomic number, where corrections for background, yield, and particle size absorption are large. However, several verifications by gravimetry and ion chromatography (Stevens *et al.*, 1978, 1980; Shaw *et al.*, 1980) add credibility to the NSI results for fine-fraction sulfur; the average of LBL sulfur results is 26% higher.

In Table 6, the January 1982 INAA results obtained at the University of Illinois are compared with the October 1981 XRF results by NSI. Good precision is obtained for most elements. However, the poor agreement between potassium and bromine results is not understood and is inconsistent with the good agreement between results by LBL and NSI.

Table 7 shows the results of XRD analyses of the coarse fractions of the 18 Teflon filters by LBL and by SDSM. The differences for crystalline composition are remarkably small considering that two different data reduction schemes were used and that the filter mass loadings were less than optimum for XRD. LBL's mass calibration technique is based on measuring standards of known mass with the same diffractometer. The LBL results are therefore absolute (in this case, converted from micrograms per cubic meter on the basis of the volume of air filtered). The total mass measured by XRD is normally less than the gravimetric mass, since amorphous substances, compounds with weak powder diffraction intensities, and compounds present at less than the 5 $\mu g\ cm^{-2}$ detection limit are not included in the results.

The LBL system has a different angular sensitivity than a conventional X-ray diffractometer with θ compensating slits. In the LBL system, at angles of < 14°, the sample coverage is reduced and the background is increased. The LBL analysis program automatically compensates for this effect, since the standards are run under the same conditions. In general, lines below 14° are not used for quantitative estimates but are used only as confirming lines in the search–match procedure.

Unfortunately, analysis of SDSM of the spectra taken with the LBL instrument was complicated by this reduced sensitivity at low angles. A number of standardized samples were scanned by both the LBL and SDSM systems. On the basis of those analyses,

Table 6. Comparison of Teflon filter INAA results obtained by the University of Illinois in January 1982 (Y) with XRF results obtained by NSI in October 1981 (X)

	$\bar{X}$ (ng m^{-3})	$\bar{Y}$ (ng m^{-3})	$\bar{Y}/\bar{X}$	A (ng m^{-3})	B	R	N
			Fine fraction				
Al	119 ± 60	117	0.98	117 (86)	−0.01 (0.58)	−0.01	17
K	129 ± 6	178	1.39	179 (103)	−0.00 (0.75)	−0.00	14
Ti	−10 ± 9	14	—	—	—	—	4
V	−2 ± 5	2	—	—	—	—	17
Mn	14 ± 2	12	0.86	1 (2)	0.77 (0.08)	0.98	17
Cu	28 ± 3	14	0.51	7 (11)	0.27 (0.38)	0.44	10
As	2 ± 3	4	1.75	3 (2)	0.68 (0.26)	0.81	15
Br	56 ± 3	98	1.76	−31 (35)	2.33 (0.59)	0.89	17
Sb	12 ± 3	7	0.60	3 (8)	0.37 (0.28)	0.64	11
Ba	0 ± 13	0	—	—	—	—	0
			Coarse fraction				
Al	1099 ± 366	654	0.60	58 (344)	0.54 (0.29)	0.69	16
K	182 ± 20	310	1.70	−366 (419)	3.71 (2.23)	0.74	11
Ti	37 ± 12	47	1.29	32 (19)	0.43 (0.41)	0.49	15
V	−5 ± 6	3	—	—	—	—	17
Mn	21 ± 3	17	0.82	1 (3)	0.78 (0.13)	0.95	17
Cu	21 ± 3	28	1.35	15 (34)	0.65 (1.50)	0.30	9
As	1 ± 2	2	—	—	—	—	6
Br	37 ± 3	326	8.90	−228 (917)	15.14 (−)	0.31	17
Sb	2 ± 3	6	—	—	—	—	3
Ba	152 ± 16	117	0.77	18 (25)	0.65 (0.11)	0.99	6

• $\bar{Y}$, $\bar{X}$ and their combined random and systematic uncertainties are average values for the N samples compared. R is the correlation coefficient, and A and B are determined by linear regression for the equation $Y = A + BX$. Results are shown whenever $N > 5$, $X > 1.5$ ng m^{-3} and $Y > 1.5$ ng m^{-3}.

Table 7.

Date/time	Lab	Coarse fraction (μg m^{-3})							
		Calcite	Quartz	Mascagnite	Dolomite	Halite	Carbon*	Other†	Total mass*
10/9/0700	LBL	5.7	5.8	2.2	1.4	—	3.9	—	38.4
	SDSM	16.6	14.9	2.9	—	—	3.9	—	
10/9/1900	LBL	6.1	4.5	2.6	2.1	—	3.5	0.2Ps	22.6
	SDSM	5.4	7.3	4.3	2.2	—	3.5	—	
11/9/0700	LBL	4.7	4.4	3.7	2.5	—	2.9	—	23.2
	SDSM	4.3	7.1	5.0	3.9	—	2.9	—	
11/9/1900	LBL	3.0	2.1	3.0	—	—	2.3	0.8Mg, 0.7Zn	13.7
	SDSM	2.9	3.1	4.8	0.6	—	2.3	—	
12/9/0700	LBL	4.0	3.2	4.7	—	—	2.5	—	17.6
	SDSM	5.2	4.8	5.1	—	—	2.5	—	
12/9/1900	LBL	5.5	3.0	2.7	1.3	—	3.1	—	18.2
	SDSM	4.0	3.2	6.4	1.5	—	3.1	—	
13/9/0700	LBL	8.8	4.8	3.6	—	—	3.5	0.7, 0.1A1	34.3
	SDSM	11.9	11.7	7.2	—	—	3.5	—	
13/9/1900	LBL	5.7	1.7	1.7	—	—	2.0	—	9.6
	SDSM	4.2	1.6	1.9	—	—	2.0	—	
14/9/0700	LBL	5.2	2.0	2.1	—	—	2.5	—	21.2
	SDSM	11.1	5.2	2.4	—	—	2.5	—	
14/9/1900	LBL	8.5	1.6	0.2	0.5	—	2.4	0.3He	14.5
	SDSM	2.2	3.2	2.4	4.4	—	2.4	—	
15/9/0700	LBL	9.2	5.1	1.1	1.5	—	3.8	0.5Ps	29.6
	SDSM	14.9	10.9	—	—	—	3.8	—	
15/9/1900	LBL	10.0	5.6	1.6	—	—	3.8	0.3Al	29.8
	SDSM	12.6	9.6	2.7	1.2	—	3.8	—	
16/9/0700	LBL	9.2	4.5	1.7	0.4	—	3.6	—	34.8
	SDSM	18.2	10.9	2.1	—	—	3.6	—	
16/9/1900	LBL	6.4	1.2	0.2	—	0.4	2.4	1.2T, 5.4Zn, 0.7Ah	11.9
	SDSM	4.7	0.8	—	1.8	2.3	2.4	—	
17/9/0700	LBL	10.5	3.9	2.0	0.9	—	2.8	1.6, 0.2Ps, 3.7Si	30.3
	SDSM	17.9	9.6	—	—	—	2.8	0.1Mi	
17/9/1900	LBL	12.9	4.3	—	4.0	—	3.5	2.2Ty, 0.8Ps	24.7
	SDSM	11.3	4.1	—	2.0	3.8	3.5	—	
18/9/0700	LBL	11.5	4.2	0.2	—	—	3.6	—	26.8
	SDSM	10.9	5.4	—	1.6	5.4	3.6	—	
18/9/1900	LBL	13.5	4.3	0.3	3.4	—	4.4	0.4Ru	45.3
	SDSM	24.7	9.1	—	7.0	—	4.4	—	
Average	LBL	7.8	3.7	1.9	1.0	0.02	3.1	1.2	24.8
	SDSM	10.2	6.8	2.6	1.5	0.6	3.1	—	

* Not measured by XRD; provided by EPA.

† Mg = magnetite; Ps = lead ammonium sulfate; Zn = zincite (ZnO); Al = anglesite ($PbSO_4$); He = hematite (Fe_2O_3); Ah = anhydrite ($CaSO_4$); Si = siderite ($FeCO_3$); Ru = rutile (TiO_2); Ty = tridymite (SiO_2).

correction factors were applied to the Houston data for the recognizable spectra below 20° 2Θ. But small quantities of such minerals as montmorillonite ($[Al,Mg]_8[Si_4O_{10}]_3[OH]_{10}\cdot 12H_2O$), muscovite, kaolinite ($Al_2Si_2O_5[OH]_4$), lead ammonium sulfate ($[NH_4]_2Pb[SO_4]_2$), and the chlorite ($Mg_3[Si_4O_{10}][OH]_2Mg_3[OH]_6$) would not be detected under the conditions of the low mass loadings on the filters. Therefore, regardless of the use of corrections at these angles, it is possible that intensities of some of those minor components are greatly in error or missing entirely in the analysis.

In the absence of data from additional prior blank-filter XRT measurements, the mass used in the normalization of the SDSM results was the total mass minus the carbon mass reported by NSI. Thus, in Table 7, all SDSM values sum to the measured mass on the filter (recalculated to micrograms per cubic meter).

In both LBL and SDSM analyses of the same raw data set for the fine fraction, the only major component recognized was $(NH_4)_2SO_4$ (mascagnite), with small amounts of $(NH_4)_2Pb(SO_4)_2$ (lead ammonium sulfate) present also. An interesting daily variation occurred in the mascagnite-to-total-mass ratio. For 10–17 September, the ratio during the 0700–1700 period was > 0.35, whereas at night-time (1700–0700) the ratio was ~0.20. Other very minor components were also found in the LBL computer-assisted spectral analysis: $(NH_4)_2\cdot Fe(SO_4)_2\cdot 6H_2O$, $ZnSO_4$ and $KAlSi_3O_8$ (microcline) in the sample collected 0700 13 September.

Comparisons of results for Nuclepore filters

NSI conducted replicate XRF measurements of the elemental content of the aerosol on the Nuclepore filters in October 1980 and September 1981. Results for aluminum, silicon and bromine were not reported because there were silicon and bromine impurities in the filter medium; barium was not reported because of a blank subtraction problem in the spectral analysis. With the exception of copper and chlorine in the fine fraction and manganese in the coarse fraction, the results agreed within the reported uncertainties, and results for the more abundant species were highly correlated.

Data on major crustal elements may be used to indicate particle loss during handling. Decreases in the average concentrations remeasured in September 1981 for the coarse (fine) fraction are as follows: 5 % (0 %) for calcium, 7 % (2 %) for iron and 2 % (1 %) for lead. Thus, an average loss of 2–6 % for major species in the coarse fraction is apparent from bulk chemical analyses. Losses as large as 25 % occurred on the first few samples. Losses for the remaining filters were much smaller.

No other cross comparisons were made on the Nuclepore filters other than in microscopy of one filter, as reported in Table 4. To compensate for the absence of data on aluminum, silicon, bromine and barium for particles on Nuclepore filters, recipients of those filters were asked to use elemental concentrations measured concurrently on Teflon filters. Table 8 compares the elemental concentrations measured on Teflon and Nuclepore filters. Because the sampler failed for the 12-h period beginning at 1900 on 18 September, only 17 12-h periods are compared. For cases in which two Nuclepore filters were used for a 12-h period, the results have been combined on a time-weighted basis to represent the entire period. The results obtained on Teflon and Nuclepore filters are nearly equivalent. For major species in each fraction, the means agree within 10 % and the correlation coefficients are quite high.

Comparison of bulk analysis with aerosol 'speciation' techniques

A direct comparison of bulk analytical methods and 'speciation' methods (such as XRD, ALM and SAX) requires that the results be reduced to a common basis. In most cases, this common ground is the bulk elemental composition measured directly by INAA or XRF, calculated from average crystalline component composition in the case of XRD, or calculated from assumed particle class molecular composition, specific gravity, and volume in the case of individual particle analyses (Johnson *et al.*, 1981). Table 9 compares the results obtained by XRD (SDSM reference-intensity method) and XRF plus IC for two of the Houston aerosol samples.

Few, if any, such cross-method comparisons have been reported. Our attempts in this regard are encouraging; the purpose is not to produce a bulk analysis from the results of a 'speciation' method, but rather to find a common comparative ground that will provide a mechanism for interfacing the various analytical methods employed in apportionment studies. Interfacing the various analytical methods may better utilize the strengths of each method and increase source resolution. For the comparisons presented in Tables 7 and 9, the amorphous components of the aerosol were not measured by XRD, but were determined by other means. A complete comparison would have to include a qualitative determination of the amorphous components that could then be partitioned by XRD through comparison of the measured versus the calculated X-ray mass absorption coefficients. Such a procedure would involve the interfacing of three analytical procedures for optimum characterization or speciation of aerosol samples: XRF for bulk elemental composition, XRD for quantitative determination of the crystalline components present, and SAX for identification and enumeration of the mass of amorphous phases. ALM could also be incorporated to determine particle compositions not deduced from SAX measures. Given optimum sample collection, the differences evident in Table 9 could be reduced by such a combination of techniques.

Table 10 compares the results of three analytical procedures applied to the three 'source samples' from the Houston study. For the XRF results, only el-

Table 8. Comparison of XRF results obtained by NSI in September 1981 for Nuclepore filters (Y) with those obtained by NSI in October 1981 for Teflon filters (X)*

	X (ng m^{-3})	Y (ng m^{-3})	$\bar{Y}/\bar{X}$	A (ng m^{-3})	B	R
			Fine fraction			
S	5080 ± 279	4759 ± 275	0.94	17 (225)	0.93 (0.04)	1.00
Cl	33 ± 6	17 ± 7	0.53	11 (31)	0.19 (0.93)	0.11
K	122 ± 6	110 ± 7	0.90	−8 (11)	0.97 (0.08)	0.99
Ca	150 ± 8	127 ± 7	0.85	−9 (16)	0.91 (0.10)	0.98
Ti	−8 ± 9	−46 ± 13	—	—	—	—
Cr	4 ± 3	3 ± 3	0.97	1 (3)	0.64 (0.53)	0.55
Mn	14 ± 2	20 ± 3	1.40	7 (3)	0.88 (0.13)	0.96
Fe	166 ± 9	146 ± 9	0.88	−14 (8)	0.97 (0.04)	1.00
Ni	4 ± 1	1 ± 2	—	—	—	—
Cu	28 ± 3	16 ± 4	0.58	3 (17)	0.46 (0.57)	0.40
Zn	82 ± 5	72 ± 5	0.88	−7 (6)	0.96 (0.05)	0.99
Sr	1 ± 1	2 ± 1	—	—	—	—
Pb	480 ± 23	464 ± 23	0.97	35 (31)	0.89 (0.06)	0.99
			Coarse fraction			
S	407 ± 430	441 ± 441	1.08	−53 (139)	1.21 (0.33)	0.90
Cl	346 ± 111	299 ± 93	0.87	−98 (32)	1.15 (0.07)	0.99
K	170 ± 20	157 ± 20	0.93	−3 (53)	0.94 (0.30)	0.86
Ca	2664 ± 144	2427 ± 129	0.91	−120 (508)	0.96 (0.18)	0.95
Ti	35 ± 12	16 ± 16	0.44	0 (16)	0.43 (0.37)	0.54
Cr	7 ± 3	11 ± 4	1.51	2 (7)	1.20 (0.82)	0.62
Mn	21 ± 3	14 ± 3	0.67	−2 (6)	0.78 (0.23)	0.88
Fe	604 ± 34	518 ± 30	0.86	−13 (93)	0.88 (0.14)	0.96
Ni	4 ± 1	4 ± 2	1.00	−0 (1)	1.04 (0.23)	0.93
Cu	18 ± 3	14 ± 4	0.80	6 (11)	0.46 (0.56)	0.41
Zn	55 ± 4	48 ± 5	0.88	−9 (7)	1.03 (0.09)	0.99
Sr	8 ± 1	7 ± 1	0.93	0 (2)	0.88 (0.18)	0.94
Pb	127 ± 10	125 ± 10	0.99	13 (10)	0.88 (0.08)	0.99

* Samples were obtained from 0700 10 September to 1900 18 September. The comparisons are based on 17 paired samples in each size fraction. $\bar{Y}/\bar{X}$ is the ratio of means, R is the correlation coefficient, and A and B are determined by linear regression for the equation $Y = A + BX$. Results are shown whenever $X > 1.5$ ng m^{-3} and $Y > 1.5$ ng m^{-3}.

Table 9. Results obtained by XRD (SDSM reference-intensity method) and by XRF, IC for two samples

	Coarse fraction (μg m^{-3})			
Element	0700 10 September		0700 13 September	
or ion	XRD	XRF, IC	XRD	XRF, IC
Si	7.0	4.9	5.5	4.6
Al	—	1.7	—	1.7
Na	—	0.1	—	—
K	—	0.2	—	0.2
Ca	6.4	3.8	4.6	3.9
Fe	0.1	1.3	0.1	1.1
Mn	0.1	0.1	0.1	0.04
Ti	—	0.1	—	0.1
Pb	—	0.1	—	0.2
Sr	—	0.02	—	0.01
SO_4^{2-}	0.7	1.1	3.0	1.7
NH_4^+	0.3	—	1.1	—

* Results are corrected for 10% fine contamination in coarse fraction.

emental ratios to silicon were reported. The elemental weight percent contributions for 12 elements, as deduced from XRF, were calculated using the XRD (SDSM reference-intensity method) value for silicon. Since these 'source samples' contained substantial amounts of quartz, and the uncertainty in the reference intensity for quartz is small, the XRD silicon value was felt to be accurate. For major elements, the three procedures seem to compare satisfactorily. This indicates a common basis for intercomparison among

Table 10. Results obtained by SAX, XRD and XRF on three Houston soil and road dust samples*

	Shell road† (UHS-1)			Soil‡ (UHS-2)			Wheeler Road§ (UHS-3)		
Element	SAX	XRD	XRF	SAX	XRD	XRF	SAX	XRD	XRF
Na	0.02	0.17	—	0.02	0.67	—	0.03	0.5	—
Mg	0.02	0.20	—	0.05	0.24	—	0.11	0.6	—
Al	2.36	1.12	3.20	4.8	3.8	6.8	2.7	3.5	6.0
Si‖	15.8	12.9	12.9	22.7	25.1	25.1	17.1	25.4	25.4
P	0.01	<0.01	0.64	0.06	—	0.4	0.15	—	0.8
S	0.29	<0.01	0.64	0.03	<0.01	0.34	0.25	<0.01	1.2
K	0.50	0.09	0.85	0.77	0.50	1.76	0.5	0.6	1.8
Ca	23.0	26.7	38.7	14.6	12.9	15.1	21	13	30
Ti	0.14	<0.01	0.60	0.51	<0.01	0.85	0.15	<0.01	0.9
Mn	0.13	0.55	0.15	0.12	0.31	0.12	0.04	0.27	0.2
Fe	1.64	0.74	3.51	3.4	0.9	5.7	3.0	0.8	7.1
Cu	0.08	—	0.10	0.04	—	0.5	0.1	—	0.6
Pb	0.32	—	0.09	0.10	—	0.08	0.13	—	—

* Elemental weight percent of the inorganic fraction.

† Dust from an unpaved road constructed of ground seashells ($CaCO_3$) at the south end of the University of Houston campus.

‡ Dust from a parking lot near the University of Houston library.

§ Dust from a paved road near the south end of the University of Houston campus.

‖ For Si, the XRF values are set equal to the XRD results; see text.

Table 11. Selected elemental ratios* calculated for two case days

	Ca/Si		Al/Si		Fe/Si		Mg/Si		S/Si	
Method	A†	B‡	A	B	A	B	A	B	A	B
SAX	1.06	2.02	0.13	0.25	0.06	0.36	0.03	0.01	0.10	0.15
XRD (SDSM)	0.92	0.84	0	0	0.02	0.02	0	0.003	0.10	0.32
XRD (LBL)	0.97	2.12	0	0	0.12	0.56	0.01	0.15	0.38	0.94
XRF	0.78	0.84	0.35	0.37	0.27	0.24	—	—	0.22	0.14

* Coarse fraction.

† A = 0700 10 September.

‡ B = 0700 13 September.

Table 12. Selected elemental ratios* averaged for the 18-filter set

Method	Ca/Si	Al/Si	Fe/Si	Mg/Si	S/Si
SAX	1.77	0.19	0.30	0.02	0.10
XRD (SDSM)	1.52	0.04	0.04	0.08	0.19
XRD (LBL)	2.18	0.02	0.03	0.06	0.56
XRF	0.93	0.37	0.20	—	0.13

* Coarse fraction.

analyses for bulk elemental composition and the several speciation methods, although refinements in the process are needed.

A 'reduction to elements' was also performed in order to compare the SAX, XRF, and XRD results obtained on the Houston aerosol samples. Tables 11 and 12 show these results for two selected filters and for the 18-filter average (coarse fraction) reported as ratios to silicon. The disagreement most obvious in these tables is the low aluminum/silicon values reported by XRD. The explanation is readily seen in the low sensitivity of XRD to aluminum-bearing clays with the light loadings characteristic of these filters.

DISCUSSION

The applicability of XRF in the quantitative analysis of aerosol samples has been discussed extensively (Dzubay and Rickel, 1978), but little has been described for XRD. The adaption of XRD to this application requires an understanding of the relative errors associated with various steps in the calibration and data reduction procedures. For absolute calibration methods, such as that employed by LBL, the uncertainties are dominated by those associated with the spectral stripping procedures and the calibration constant measurements. The reference-intensity method has an additional uncertainty associated with the determination of the mass of amorphous materials present.

If these errors are describable by normal statistical methods, their influence on the final determination can be incorporated into the final apportionment. Possible systematic effects include artifacts in the spectral stripping results caused by overlapping peaks and possible variations in calibration coefficients caused by anomalous scattering intensities observed for certain mineral specimens. These systematic effects are closely monitored and corrections made where appropriate. A method for estimating final component uncertainties has been developed at SDSM (Davis, 1981b). This approach provides a practical means to propagate individual measurement errors. In the present analyses of the Houston filters, directly measuring the sample mass absorption coefficient (μ_{s0}^{*}) was not possible because the original tared filters were not submitted to the SDSM laboratories for XRT measurements. As a result, μ_{s0}^{*} had to be estimated qualitatively on the basis of the observed components present, and an additional unknown error arose from this source that was not included in the error propagation procedure. Iterative tests on several analyses have shown the error in intensity corrections resulting from this factor to be small, less than 5%, so that no serious additional uncertainties should arise from this cause. Lack of the XRT data, however, also prohibited measurement of the amorphous component on the Houston filters. A similar error propagation analysis was completed on the soil and road dust samples; in these cases, the additional XRT data were available and the μ_{s0}^{*} uncertainty not applicable.

Overall, the comparability of results obtained by various analytical techniques in this exercise seems quite good. None of the samples compared were optimum, in terms of substrate or mass loading, for more than one technique. Suitable standards need to be developed, and a sampling protocol for ambient aerosol needs to be established to optimize conditions for several techniques. But regardless of the comparison accuracy between analytical methods, source apportionment results based upon trace substance quantification will be internally consistent as long as both source and ambient samples are characterized by the same method at the same facility.

The source apportionment process using XRD can be carried out independently of other analytical methods under appropriate sampling and filter processing conditions (Davis *et al.*, 1981, in press). However, in a field and laboratory 'after-the-fact' study not initially designed to accommodate XRD analysis (as in the present case), many limitations of this approach become apparent. These include:

(1) For dichotomous sampling, the coarse-fraction mass per unit area on the filter should be $\sim 200\ \mu g\ cm^{-2}$. For conventional systems, the loading should be $> 200\ \mu g\ cm^{-2}$ for measurement of minor components. Where masses of $> 300\ \mu g\ cm^{-2}$ are anticipated, fused quartz filters provide a great advantage with these systems through the elimination of substrate peak interference. The improved sensitivity of the LBL instrument permits loadings nearer to the usual values for field samples ($\sim 100\ \mu g\ cm^{-2}$), but the quantification of components present at $< 5\ \mu g\ cm^{-2}$ is still difficult. Quartz filters are not needed with this system.

(2) Quantitative analysis of samples containing many components is made difficult by the overlapping of spectral profiles.

(3) Direct quantitative measurement of amorphous components as well as of crystalline components by the SDSM XRD/XRT method requires preliminary XRT measurements of blank filters prior to loading.

(4) Source sampling must be carefully considered, and on-site point and area source sampling conducted where possible. Inasmuch as the bulk of the coarse-fraction mass comes from fugitive dust sources, proper soil and street particulate sampling is especially important. Soil and road dust sampling should include compositing of separate samples in the vicinity of the receptor site to insure representativeness of this fugitive dust source. Where possible, parent mineral materials for the region under study should be obtained, since considerable variation exists in the reference intensity constant (or mass calibration curve) for a given mineral on a geographic basis. This would be a key element in reducing the uncertainty in the SDSM reference-intensity method as well as providing an appropriate set of standards for the LBL mass calibration.

The amorphous component determination using XRD/XRT and supported by ALM and SAX constitutes a potentially valuable approach. Although this phase of measurement remains partly in the developmental stage, we believe it is a potentially useful approach to source apportionment studies. The resolution of any receptor model for source apportionment is directly related to the 'degrees of freedom' in the data partitioning (*N*-1 observation variables). Using microscopic techniques to identify amorphous components serves to enhance the crystalline compound results. Similarly, XRD data can be used in SAX particle classification schemes to improve the accuracy of particle type identifications.

SAX, like XRD, can be applied independently of other analytical methods as the basis for source apportionment (Johnson and McIntyre, 1983; Johnson and Twist, 1983), though it also has a strong supportive role for more established methods. The chief advantage of SAX is the speed with which individual particles can be characterized. In such an analysis, the information contained in the heterogeneous nature of the sample is preserved. This offers a powerful contrast to bulk analytical methods. The creation and storage of a discrete object vector for each particle facilitates data reduction by a variety of mathematical techniques (factor analysis, discriminant analysis, and linear regression). The use of size and type categories can reduce or eliminate elemental signature collinearities in source samples. Size may be presented on either the physical (measured) or the estimated aerodynamic (mass median diameter) basis, and may include, with some SAX systems, area-to-perimeter ratios for the particle images.

Of course, some analytical limitations are also attendant to SAX: samples must be compatible with high vacuum, particle classifications are not always phase- and/or compound-specific, and mass load on the filter is restricted to a monolayer of particles with $< 10\%$ overlapping and with an average separation of 1.5 times the diameter.

SAX methods of apportionment are subject to the same constraints and assumptions that apply to other receptor models. The results of analyses of ambient samples must be apportioned to known (characterized) sources by one or more mathematical techniques. In this respect, the computer-assisted scanning electron microscopic methods follow traditional pathways. However, they are capable of greater source resolution than some other methods. All receptor models assume that emitted materials obtained at the receptor site can be represented by a linear combination of fractional contributions from the variety of potential sources. As one attempts to increase source resolution from generic sources to specific sources, this assumption becomes more questionable. Such a concern is particularly true for individual particle characterization, since gas-to-particle conversions and particle aggregation phenomena may affect the identity of particles during transit from source to site. All receptor models are sensitive to variations in source strength and emissions character. Since small changes in the physical source processes may affect individual particle morphology and chemical makeup on a *particle-by-particle* basis more than on a bulk chemical composition basis, SAX procedures are vulnerable.

A trained analytical light microscopist can identify source types from particle features without any additional source information, but the optical criteria can only be optimized with reliable source samples and data. The technique has powerful applications in receptor modeling, but may be limited by cost or availability of expertise. Further, particles of $< 1\ \mu m$ are below the effective resolution of ALM. Such small particles can be seen, but the amount of useful analytical data collectable is limited.

Since the ALM classification of particles into categories is based on optical criteria, particles that do not exhibit many optical properties are less specifically identified. Samples that contain soot, coal, coke, oil coatings, iron scale, charred biologicals and other black opaque materials result in a reduction in the reliability of the assignment of any individual particle. But ALM is sensitive to crystalline, noncrystalline, combusted, biological, and mixed-phase particles in the coarse mode. It is uniquely appropriate to the analysis of particulate samples involving particles of $> 1\ \mu m$.

ALM, XRD and SAX are complementary. Each technique has capabilities beyond the range of the others and complementary to them. ALM can often be used to resolve sources that may be indistinguishable using other techniques. Similarly, the opaque particulate material that cannot be well distinguished by ALM can often be better resolved using SAX or XRD.

CONCLUSIONS

The Quail Roost II exercises led to several major conclusions about ambient and source sampling and analysis for purposes of source apportionment:

(1) The resolution of receptor models can be improved by the inclusion of several analytical techniques for the chemical partitioning of aerosol samples. Bulk, crystalline compound, and particle type characterizations are complimentary.

(2) To achieve suitable analyses of the same aerosol specimen by several techniques requires certain compromises in sample collection procedures. With careful preplanning, however, analysis of the same specimen by more than one technique is possible.

(3) There does appear to be a common ground for the comparison of results obtained by several analytical techniques. Particle type and crystalline component information can be reduced to elemental composition for comparison with XRF or INAA bulk elemental results. Future comparisons of this nature should include analyses of appropriate NBS (or other) standard reference materials. An analytical cross-comparison should preceed any results based upon ambient samples.

(4) Microscopic techniques and crystalline component determinations are sufficiently descriptive and quantitative to serve as the basis, in the absence of additional analyses, for atmospheric particle source apportionment. Their greatest power, however, is in a combined receptor model, where they can be focused on areas of poor source resolution.

REFERENCES

Courtney W. J., Shaw R. W. and Dzubay T. G. (1982) Precision and accuracy of a beta-gauge for aerosol mass determinations. *Envir. Sci. Technol.* 16, 236–239.

Davis B. L. (1978) Additional suggestions for X-ray quantitative analysis of high-volume filter samples. *Atmospheric Environment* 12, 2403–2406.

Davis B. L. (1980) "Standardless" X-ray diffraction quantitative analysis. *Atmospheric Environment* 14, 217–220.

Davis B. L. (1981a) The use of X-ray diffraction quantitative analysis in air quality source studies. In *Electron Microscopy and X-Ray Applications to Environmental and Occupational Health Analysis*, Vol. 2 (Edited by Russell), pp. 131–152. Ann Arbor, Ann Arbor, Michigan.

Davis B. L. (1981b) A study of the errors in X-ray quantitative analysis procedures for aerosols collected on filter media. *Atmospheric Environment* 15, 291–296.

Davis B. L. (1982) Hybrid model for source apportionment of the Houston aerosol for the period 10–18 September 1980.

Reproduced from
best available copy.

Report 82-3. Institute of Atmospheric Sciences, South Dakota School of Mines and Technology, Rapid City, South Dakota.

Davis B. L., Johnson L. R. and Flannagan M. J. (1981) Provenance factor analysis of fugitive dust produced in Rapid City, South Dakota. *J. Air Pollut. Control Ass.* **31,** 241–246.

Davis B. L. Maughan D. and Carlson J. (1983) X-ray studies of airborne particulate matter observed during wintertime at Missoula, Montana. In *Electron Microscopy and X-Ray Applications to Environmental and Occupational Health Analysis*, Vol. 3 (edited by Russell), pp. 131–146. Ann Arbor, Ann Arbor, Michigan.

Deer W. A., Howie R. A. and Zussman J. (1962) *Rock Forming Minerals. Vol. 5 non-Silicates.* John Wiley, New York.

Drane E. A., Rickel D. G., Courtney W. J. and Dzubay T. G. (1980) Computer code for analyzing X-ray fluorescence spectra on airborne particulate matter. In *Advances in X-Ray Analysis*, Vol. 23 (edited by Rhodes), pp. 149–156. Plenum Publishing Corporation, New York.

Dzubay T. G. (1977a) *X-Ray Fluorescence Analysis of Environmental Samples.* Ann Arbor, Ann Arbor, Michigan.

Dzubay T. G. (1977b) Recent advances in energy-dispersive XRF analysis of aerosol samples. *Trans. Am. Nucl. Soc.* **27,** 203–204.

Dzubay T. G. and Nelson R. O. (1975) Self absorption corrections for X-ray fluorescence analysis of aerosols. In *Advances in X-Ray Analysis*, Vol. 18, pp. 619–631. Plenum Publishing Corporation, New York.

Dzubay T. G. and Rickel D. G. (1978) X-ray fluorescence analysis of filter-collected aerosol particles. In *Electron Microscopy and X-Ray Applications to Environmental and Occupational Health Analysis* (edited by Russell and Hutchings), pp. 3–20. Ann Arbor, Ann Arbor, Michigan.

Dzubay T. G., Stevens R. K., Lewis C. W., Hern D., Courtney W. J., Tesch J. W. and Mason M. A. (1982) Visibility and aerosol composition in Houston, Texas. *Envir. Sci. Technol.* **16,** 514–525.

Hanna R. B., Karcich K. J., and Johnson D. L. (1980) Determination of particle identities via a computer assisted SEM-EDAX system. *Scanning Electron Microsc.* **1,** 323–327.

Jaklevic J. M., Landis D. A. and Goulding F. S. (1976) Energy dispersive X-ray fluorescence spectrometry using pulsed X-ray excitation. In *Advances in X-Ray Analysis*, Vol. 19 (edited by Gould *et al.*), pp. 253–265. Kendall Hunt, Dubuque, Iowa.

Jaklevic J. M., Gatti R. C., Goulding F. S. and Loo B. W. (1981) A beta-gauge method applied to aerosol samples. *Envir. Sci. Technol.* **15,** 680–686.

Johnson D. L. (1983) Automated scanning electron microscopic characterization of particulate inclusions in biological tissues. *Scanning Electron Microsc.* **III,** 1211–1228.

Johnson D. L. and McIntyre B. L. (1983) A particle class balance receptor model for aerosol apportionment in Syracuse, NY. In *Receptor Models Applied to Contemporary Pollution Problems, Proc. APCA Specialty Conf.* SP-48 (Danvers, Massachusetts, October 1982), pp. 238–248. Air Pollution Control Association.

Johnson D. L. and Twist J. P. (1983) Statistical considerations in the employment of SAX results for receptor models. In *Receptor Models Applied to Contemporary Pollution Problems, Proc. APCA Specialty Conf.* SP-48 (Danvers, Massachusetts, October 1982), pp. 224–237. Air Pollution Control Association.

Johnson D. L., McIntyre B., Fortmann R., Stevens R. K. and Hanna R. B. (1981) A chemical element comparison of individual particle analysis and bulk chemical analysis. *Scanning Electron Microsc.* **1,** 469–476.

Lee R. J. and Kelly J. F. (1980) Overview of SEM-based image analysis. *Scanning Electron Microsc.* **1,** 303–310.

Maney J. P., Fasching J. L. and Hopke P. K. (1977) A versatile and comprehensive analysis code for automated reduction of gamma-ray spectral data. *Computers Chem.* **1,** 257–264.

O'Connor B. H., Kerrigan G. C., Thomas W. W. and Pearce A. T. (1977) Use of bromine levels in airborne particulate samples to infer vehicular lead concentrations in the atmosphere. *Atmospheric Environment* **11,** 635–638.

Roscoe B. A. (1981) Applicability of factor analysis to the determination of mineral matter in coal. Ph.D. dissertation, University of Illinois, Urbana, Illinois.

Shaw R. W., Stevens R. K. and Courtney W. J. (1980) Exchange of comments on inaccuracies encountered in sulfur determination by particle induced X-ray emission. *Analyt. Chem.* **52,** 2217–2218.

Stevens R. K. and Pace T. G. (1984) Overview of the mathematical and empirical receptor Models Workshop (Quail Roost II). *Atmospheric Environment* **18,** 1499–1506.

Stevens R. K., Dzubay T. G., Russwurm G. and Rickel D. (1978) Sampling and analysis of atmospheric sulfates and related species. *Atmospheric Environment* **12,** 55–68.

Stevens R. K., Paur R. J., Lewis C. W. and Dzubay T. G. (1979) Houston aerosol characterization: EPA measurements. In *Proceedings, Ozone/Oxidants Interactions with the Total Environment* (edited by Fredrick), pp. 212–220. Air Pollution Control Association, Pittsburgh, Pennsylvania.

Stevens R. K., Dzubay T. G., Shaw R. W., McClenny W. A., Lewis C. W. and Wilson W. E. (1980) Characterization of the aerosol in the Great Smoky Mountains. *Envir. Sci. Technol.* **14,** 1491–1498.

Stevens R. K., McClenny W. A., Dzubay T. G., Mason M. A. and Courtney W. J. (1982) Analytical methods to measure the carbonaceous content of aerosols. In *Particulate Carbon Atmospheric Life Cycle* (edited by Wolff and Klimisch), pp. 111–126. Plenum Press, New York.

Thompson A. C., Jaklevic J. M., O'Conner B. H. and Morris C. M. (1982) X-ray powder diffraction system for chemical speciation of particulate aerosol samples. *Nucl. Instrum. Methods Phys. Res.* **198,** 539–546.

Atmospheric Environment Vol. 18, No. 8, pp. 1555–1566, 1984
Printed in Great Britain.

INTERLABORATORY COMPARISON OF RECEPTOR MODEL RESULTS FOR HOUSTON AEROSOL*

THOMAS G. DZUBAY,† ROBERT K. STEVENS
U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, U.S.A.

W. DAVID BALFOUR, HUGH J. WILLIAMSON
Radian Corporation, Austin, TX 78766, U.S.A.

JOHN A. COOPER, JOHN E. CORE, RICHARD T. DE CESAR
NEA, Inc. Beaverton, OR 97005, U.S.A.

E. RUSSELL CRUTCHER
Microlab Northwest, Redmond, WA 98052, U.S.A.

STUART L. DATTNER
Texas Air Control Board, Austin, TX 78723, U.S.A.

BRIANT L. DAVIS
South Dakota School of Mines and Technology, Rapid City, SD 57701, U.S.A.

STEVE L. HEISLER, JITENDRA J. SHAH
Environmental Research and Technology, Inc., Concord, MA 01742, U.S.A.

PHILIP K. HOPKE
Institute for Environmental Studies, University of Illinois at Urbana-Champaign, Urbana, IL 61801, U.S.A.

and

DAVID L. JOHNSON
Department of Chemistry, State University of New York College of Environmental Science and Forestry,
Syracuse, NY 13210, U.S.A.

(*First received 9 June* 1983 *and in final form* 19 *January* 1984)

Abstract—In exercises for the Mathematical and Empirical Receptor Models Workshop (Quail Roost II), nine investigators independently applied receptor models to apportion ambient aerosol mass concentrations to components from emission sources in Houston, TX. The receptor model results were based upon ambient aerosol compositional data supplied to investigators that included mass, elemental and ionic species concentrations for consecutive 12-h sampling periods at the University of Houston between 10 and 19 September 1980. Some investigators performed additional analyses by X-ray powder diffraction, scanning electron microscopy with automated image analysis and X-ray energy spectroscopy, and light microscopy. In most cases, the fine (0–2.5 μm) and coarse (2.5–15 μm) fractions were analyzed separately by the receptor models. The models tested include effective variance type chemical mass balance, weighted ridge regression, multiple linear regression, factor analysis and target transformation factor analysis. The number of source classes determined ranged from 4 (by multiple linear regression) to 23 (by light microscopy). Results are presented and compared in eight general emissions categories.

INTRODUCTION

Receptor models are mathematical procedures for identifying and quantifying the sources of ambient aerosol measured at a receptor site. There are several types of receptor models (Watson *et al.*, 1981; Henry *et al.*, 1984), and the comparability of results obtained by the various types had not been tested previously. At the Mathematical and Empirical Receptor Models

---

* A Report from the Mathematical and Empirical Receptor Models Workshop (Quail Roost II).

† To whom requests for reprints of this article should be addressed. Requests for reprints of the entire series of six articles (including this one) documenting the Quail Roost II workshop should be directed to Robert K. Stevens at the same address.

Workshop (Quail Roost II) (Stevens and Pace, 1984), the results of receptor models applied to two types of aerosol data were intercompared. The first type consisted of sets of 'aerosol data' simulated mathematically, and the results are described by Currie *et al.* (1984). The second type consisted of ambient aerosol data for samples collected in Houston, TX, and is the subject of this report. Several features of the Houston data set distinguish it from the simulated data sets: (1) The true impacts of emission sources in Houston are unknown. (2) The number of sources that contribute to the aerosol in Houston is unknown and must greatly exceed the 13 sources used by Currie *et al.* (1984) in the simulated data. (3) Because source signatures were not measured in Houston, signatures measured elsewhere had to be used. (4) Human judgement was required to determine the numbers and types of source signatures that were used in chemical mass balance analyses of the Houston data set.

Workshop participants applied seven different receptor models to the Houston data set. The data were supplied to the participants five weeks prior to the workshop. The ambient data were obtained by operating four dichotomous samplers simultaneously at a site on the University of Houston campus (Dzubay *et al.*, 1982). Filters were changed at 0700 and 1900 Central Daylight Time for 18 consecutive periods between 10 and 19 September 1980. The dichotomous samplers collected particles in the fine (0–2.5 μm) and coarse (2.5–15 μm aerodynamic equivalent diameter) size fractions. The filter samples were analysed for mass, elemental and ionic species concentrations (Dzubay *et al.*, 1982; Johnson *et al.*, 1984), and the concentrations in the fine and coarse fractions for each of the 18 sampling periods were sent to all participants. The average concentrations are shown in Table 3 of Johnson *et al.* (1984). Participants also received a compilation of hourly wind vectors measured at a site located 20 km north of the University of Houston.

In addition, filter samples from Houston were sent to some participants who performed analyses by X-ray powder diffraction (XRD), light microscopy (LM), and scanning electron microscopy with automated image analysis and X-ray energy spectroscopy (SAX). These methods and results are described by Johnson *et al.* (1984). The results of these additional analyses were not distributed to participants prior to the workshop, and only those who performed the analyses were able to use the resulting data in their receptor model calculations.

Participants were sent two different tabulations of emission rates of sources in the Houston area. One was obtained from the U.S. Environmental Protection Agency's National Emissions Data System (NEDS); the second was from a report by Heisler *et al.* (1981). Although the compositions of particles from emission sources in Houston were not measured, participants were provided with compilations of source signatures from other areas by Watson (1979) and by Cass and McRae (1981). In addition, the chemical signatures of a soil sample and two road dust samples were provided. The soil sample was collected near the University of Houston's library, and the road dust samples were taken from the south end of the campus. One of these roads was unpaved and constructed of crushed seashells, making it rich in calcium carbonate.

SOURCE APPORTIONMENT PROCEDURES AND RESULTS

The categories to be used in reporting results were not specified in advance. Consequently, the various participants used about 40 different ones. To facilitate comparisons, the participants' results were condensed into eight general categories. Table 1 shows these general categories as well as the specific ones that contribute to them. The general categories labelled "sulfate," "carbon" and "nitrate" pertain to those species and related ions in excess of the amounts included in the other five general categories. Except as noted below, participants analyzed the coarse and fine fractions separately. Six participants based their receptor model calculations on elemental and ionic compositional data; one participant used carbon and XRD data; one used SAX data; and one used LM data.

Table 2 indicates the marker species used by the six participants whose calculations were based on elemental and ionic composition, and Table 3 shows the abundances of some of the marker species used in their transportation and sulfate components. Tables 4 and 5 present the results for the eight general categories for daytime September 10 and for the entire period from 0700 10 September to 0700 19 September. Participants' original results for coarse-fraction crustal matter for daytime 10 September are detailed in Table 6. The specific procedures used by each participant are outlined below. Emphasis is placed on the transportation and sulfate categories in the fine fraction and the crustal category in the coarse fraction, since they are responsible for the bulk of the total mass concentration.

Source apportionment using ionic and elemental composition

Effective variance chemical mass balance [*CMB(EV)*]. Shah of Environmental Research and Technology, Inc. (ERT-A) used a CMB(EV) method described elsewhere (Currie *et al.*, 1984). Source profiles were selected from a library containing about 25 profiles from Cass and McRae (1981) and Watson (1979). A profile for noncatalyst car emissions and another for a weighted mixture of car and truck emissions (Shah, 1981) were included. The latter was based on the tunnel data of Pierson (1979) and the assumption that 5% of the vehicle miles travelled involved heavy-duty diesel trucks. For some samples, the transportation profile was based on the weighted mixture; for others, it was based on the emissions from noncatalyst cars alone. Solutions were obtained using six to nine profiles in each of the fine- and coarse-fraction samples. For each sample, fine-fraction sulfate was designated to be either ammonium sulfate or ammonium bisulfate (but not both).

To test for effects related to selections of marker elements and source profiles (which were subjective), a second investigator, Heisler of Environmental Research and Technology (ERT-B), independently performed analyses using a procedure similar to that of Shah. The number of source profiles used to fit a given sample ranged from 7 to 11 for the fine fraction and from 6 to 9 for the coarse fraction. For any particular sample, ERT-A and ERT-B did not always use the same set of marker elements and source profiles. For example, ERT-B always used both ammonium sulfate and ammonium bisulfate as profiles for sulfate in the fine fraction, whereas ERT-A always used only one of them. The consequences of such a difference were very small. However, the meaning of

Table 1. Source categories and marker elements used in source apportionment of Houston aerosol

Marker elements	General source category	Specific source categories
Na, Cl	Marine	Marine
Al, Si, K, Ca, Mn, Fe	Crustal	Wind erosion of soil
		Paved and unpaved roads
		Construction
		Limestone crushing and handling
		Cement
		Lime kiln
		Fly ash
		Slag
C, Br, Pb	Transportation	Noncatalyst vehicles
		Catalyst cars
		Diesels
Mn, Fe	Steel	Iron and steel production
		Steel finishing and handling
Cu, Zn, As, Sm, Sb	Nonferrous metals	
NH_4^+, SO_4^{2-}	Sulfate	Primary emissions in Houston area
		Conversion of SO_2 in Houston area
		Regional background
C	Carbon	Refineries
		Botanical
		Combustion of vegetative matter
		Tire wear
NO_3^-	Nitrate	Photochemical

Table 2. Marker species used in source apportionment of the fine (F) and coarse (C) fractions

Species	ERT-A CMB(EV)	ERT-B CMB(EV)	NEA CMB(EV)	Radian CMB(RR)	TACB MLR(T) FA	UI TTFA
C	F, C	F, C	F, C	F, C		F, C
NH_4^+	F	F, C	F			
NO_3^-	F, C	C	C	F, C		F, C
Na^+	F*, C	F, C	C			C
Al	F*, C*		F*, C	F, C		C
Si	F, C	F, C	F, C	F, C	C	C
P						F, C
S					F	
SO_4^{2-}	F, C	F, C	F, C	F, C		F, C
Cl	C	C*	C	F, C		F, C
K	F, C	F, C	C	F, C		F, C
Ca	F, C	F, C	F, C	F, C	C	F, C
Ti	F*, C*	F, C*	C*	F, C		C
V				F, C		C
Cr	C*	F, C		F, C		
Mn		F, C	F*, C	F, C		F, C
Fe	F, C	F, C	F, C	F, C		C
Ni		F, C		F, C		F
Cu	C*			F, C		F, C
Zn	C*		F, C*	F, C		F, C
As			F*	F, C		
Se				F, C		
Br	C	C	F, C	F, C	F	F, C
Rb				F, C		
Sr				F, C		C
Cd				F, C		
Sn			F*	F, C		
Sb			F*	F, C		
Ba				F, C		F, C
Pb	F, C	F, C	F, C	F, C		F, C

* Indicates species used for some but not all samples.

Table 3. Abundances of marker species in five fraction profiles (transportation and sulfate categories) used by various investigators

		Abundance (%)					
Category, profile	Species	ERT-A CMB(EV)	ERT-B CMB(EV)	NEA CMB(EV)	Radian CMB(RR)	TACB MLR(T)/FA	UI TTFA
Transportation							
Noncatalyst cars	Pb	21.1	21.1		21.1		
Weighted mixture (cars, trucks)	Pb	13.0	13.0	13.8			
MLR(T)/FA, TTFA results	Pb					4.8*	4.7
MLR(T)/FA, TTFA results	Br					0.57†	0.50
Sulfate							
Ammonium sulfate	SO_4^{2-}	72.7	72.7				
Ammonium bisulfate	SO_4^{2-}	84.5	84.5				
Sulfate ion	SO_4^{2-}			100.0			
Ammonium ion	NH_4^+			100.0			
MLR(T)/FA, TTFA results	SO_4^{2-}					50.0‡	53.8

* An estimate consisting of the reciprocal of the regression coefficient for Br multiplied by the ratio of the mean ambient Pb to mean ambient Br concentration.

† Reciprocal of the regression coefficient for Br.

‡ Three times the reciprocal of the regression coefficient for S.

Table 4. Intercomparison of source apportionment results ($\mu g\ m^{-3}$) for the 12-h period beginning at 0700 on 10 September 1980*

Size range	Category	ERT-A CMB(EV)	ERT-B CMB(EV)	NEA CMB(EV)	TACB MLR(T)/FA	UI TTFA	SDSM CMB(XRD)	SUNY SAX	MN LM
Fine	Marine	0.58	0.53						
Fine	Crustal	3.9	1.9	1.7					
Fine	Transportation	3.8	2.3	3.7	11.5	17.6	0.46		
Fine	Steel	1.9	1.8	0.74					
Fine	Nonferrous metals			0.11					
Fine	Sulfate	23.4	25.6	25.8	41.1		50.8		
Fine	Carbon	6.8	8.8	7.7		36.3			
Fine	Nitrate								
Fine	Total	40.4	40.9	39.8	52.6	53.9	51.3		
Coarse	Marine	0.26	0.23	0.27		0.33	<0.01	0.0	
Coarse	Crustal	17.0	24.3	31.4	39.9	40.3	41.2	28.9	30.6
Coarse	Transportation	0.69	0.42	0.95			0.05	0.3	
Coarse	Steel		0.12	0.33				3.1	3.0
Coarse	Nonferrous metals			0.04				0.9	
Coarse	Sulfate	1.4	1.0	1.1		2.5	1.1	0.6	4.7
Coarse	Carbon	1.1	1.6	3.1				0.3	
Coarse	Nitrate	1.3	0.57	1.0		0.57			
Coarse	Total	21.7	28.2	38.2	39.9	43.7	43.3	34.1	38.3

* The mass concentrations measured in the fine and coarse fractions were 56.3 ± 2.8 and 38.4 ± 2.8 $\mu g\ m^{-3}$, respectively.

the transportation category is ambiguous because source profiles describing a single vehicle type (noncatalyst cars) were used interchangeably with a profile describing a mixed population of cars and trucks.

De Cesar, Cooper and Core of NEA, Inc. (NEA) used a CMB(EV) method similar to that of ERT-A and ERT-B. Source profiles were selected from a library containing 33 source profiles. The library included the three road dust and soil profiles from Houston. Most of the other profiles were taken from the Portland Aerosol Characterization Study (Copper *et al.*, 1979). Sulfate and ammonium were fit with single-species profiles as shown in Table 3. The transportation profile for the fine fraction pertained to a weighted mixture of cars and trucks; the lead content is shown in Table 3. The number of profiles used in the least-squares fits ranged from 6 to 8 for the fine fraction and from 6 to 9 for the coarse fraction. In general, there were no problems in condensing the 6 to 11 categories used by ERT-A, ERT-B and NEA into the 8 general categories.

Weighted ridge regression chemical mass balance [*CMB(RR)*]. Williamson, Balfour, and co-workers at Radian Corporation applied both weighted least-squares and CMB(RR) methods to the Houston data set in manners described by Williamson *et al.* (1982), Henry *et al.* (1984) and Currie *et al.* (1984). The weighting coefficients consisted of uncertainties in the ambient concentrations. Computations were made using the average daytime and average night-time concentrations measured between 0700 11 September and 0700 19 September. The source profiles used in the analysis were taken from Watson (1979) and Cass and McRae (1981). In both the least-squares and CMB(RR) methods, 18 profiles were used; this relatively large number of profiles provided a severe test for the effects of multicollinearity.

Radian's CMB(RR) results for the 18 components were less

Table 5. Intercomparison of source apportionment results ($\mu g\ m^{-3}$) for the 9-day period beginning at 0700 on 10 September 1980*

Size rage	Category	ERT-A CMB(EV)	ERT-B CMB(EV)	NEA CMB(EV)	Radian CMB(RR)	TACB MLR(T)/FA	UI TTFA	SDSM CMB(XRD)	SUNY SAX
Fine	Marine	0.45	0.43		0.0				
Fine	Crustal	1.7	1.1	0.96	1.3				
Fine	Transportation	1.7	2.1	2.9	†	9.7	8.5	3.6	
Fine	Steel	0.58	0.49	0.15	0.22				
Fine	Nonferrous metals			0.09					
Fine	Sulfate	17.4	18.8	18.8	†	29.0	30.2	35.2	
Fine	Carbon	5.6	5.8	5.0	†				
Fine	Nitrate	0.58							
Fine	Total	28.0	28.7	27.9		38.7	38.7	38.8	
Coarse	Marine	0.99	1.0	0.9	0.88		0.7	0.8	0.25
Coarse	Crustal	16.5	18.6	17.4	27.8	27.3	20.4	24.3	18.1
Coarse	Transportation	0.65	0.41	0.89	†			0.5	0.3
Coarse	Steel		0.10	0.11	0.51				1.7
Coarse	Nonferrous metals	0.41		0.02					0.5
Coarse	Sulfate	0.79	0.53	0.81	†		1.3	1.3	0.6
Coarse	Carbon	0.35	0.52	2.4	†				0.2
Coarse	Nitrate	1.5	1.3	1.5			2.3		
Coarse	Total	21.2	22.5	24.0		27.3	24.7	26.9	21.7

* The mass concentrations measured in the fine and coarse fractions were 38.6 and 24.8 $\mu g\ m^{-3}$, respectively.

† See discussion in text.

affected by multicollinearity than their least-squares results. The concentrations predicted by CMB(RR) were positive numbers for almost every source, whereas about 35% of the concentrations predicted by the least-squares method were negative. Moreover, the predicted standard errors σ were smaller for CMB(RR) by factors of typically 2–5, and for some sources smaller by a factor of 50 or more. Despite the fact that the uncertainties were smaller for CMB(RR), the case of 18 profiles was a very severe test in terms of multicollinearity, and the reported concentrations exceeded 2σ for two sources in each size fraction.

Seven of the 18 fine-fraction components reported by CMB(RR) could be condensed into the general categories; the results are shown in Table 5. The remaining 11 fine-fraction components deduced by CMB(RR) are listed below: (Results are shown in micrograms per cubic meter, with the standard error σ in parentheses.)

slash burn [C]	0.01	(0.13)
catalyst equipped autos [C, SO_4^{2-}]	11.5	(5.2)
noncatalyst equipped autos [C]	0.27	(0.19)
jet aircraft [C]	2.9	(4.4)
residual oil [SO_4^{2-}]	0.04	(0.10)
distillate oil [C]	0.74	(0.53)
natural gas combustion [SO_4^{2-}]	0.42	(0.96)
Kraft boiler recovery [SO_4^{2-}]	0.83	(0.39)
cement production [SO_4^{2-}]	1.6	(1.1)
calcination of gypsum [SO_4^{2-}]	0.11	(0.38)
petroleum production: FCC units, CO boiler [SO_4^{2-}]	13.2	(12.9)

In this list, species whose abundances exceed 35% in the source profile are shown in brackets. Thus, it appears that the above 11 components belong in the general categories of transportation, sulfate and carbon, but it was not possible to assign them unambiguously to those categories. For catalyst and noncatalyst equipped cars, CMB(RR) determined the ratio of emission impacts to be 42 with a standard error of 35. Other, independent sources of data indicate that the ratio should be < 1. For example, only about half of the gasoline sold in the U.S. during 1980 was the unleaded type, which is used in catalyst cars (Du Pont, 1982). Also, the particulate emission rates of catalyst equipped cars are generally less than those of cars without catalysts (Hare and Black, 1981; Lang *et al.*, 1981). Thus, the CMB(RR) result for 'catalyst equipped autos' may include contributions from nonautomotive sources of carbon and sulfate. Likewise, the 'jet aircraft' component, whose profile consits of 96% carbon, may contain carbon emitted by diesel engines or several other types of carbon-emitting point sources. Finally, the CMB(RR) solution did not include profiles for sulfate from the regional background and locally produced secondary sulfate. This could have led to overestimation of the contributions from sulfate-emitting sources such as catalyst equipped autos, cement production, and petroleum production's FCC units.

Multiple linear regression by tracers and factor analysis [MLR(T)/FA]. Dattner of the Texas Air Control Board (TACB) analyzed Houston aerosol using both factor analysis and backward stepwise unweighted multiple linear regression (Kleinman *et al.*, 1980; Henry *et al.*, 1984; Currie *et al.*, 1984). Table 7 shows the factor loading matrices obtained by factor analysis after VARIMAX rotation. In addition to the components shown, sodium, nitrate, strontium, chlorine and copper were identified as single-species components in the fine fraction, as were sodium, nickel and copper in the corase fraction. The single-species components have little significance because the concentrations of those species were near the detection limits of the measurement methods.

In each size fraction, two species were determined to be significant by multiple regression analyses. In the fine fraction, those species were sulfate and bromine; in the coarse fraction, silicon and calcium. It was assumed that sulfate and bromine were representative of the sulfate and transportation categories, respectively, and that silicon and calcium together represented the general crustal category. The results are presented in Tables 4 and 5. The small number of samples in the data set prevented detection of additional components that contributed significantly to the variance in mass. At least 40 samples are needed for multivariate analysis (Henry *et al.*, 1984), but the Houston data set contained only 18.

Target transformation factor analysis (TTFA). Hopke and co-workers at the University of Illinois (UI) applied TTFA in

Table 6. Intercomparison of source apportionment results ($\mu g\ m^{-3}$) for coarse-fraction crustal matter for the 12-h period beginning at 0700 on 10 September 1980

Crustal matter	ERT-A CMB(EV)	ERT-B CMB(EV)	NEA CMB(EV)	TACB MLR(T)/FA	UI TTFA	SDSM CMB(XRD)	LBL XRD	SUNY SAX	MN LM
Soil (UHS-2, campus library)			16.7						
Road (UHS-1, shell road)	17.0	16.1				1.8			
Road (UHS-3, Wheeler Road)			14.7			} 40.3		14.5	
Construction									
Clay								3.9	
Quartz									10.6
Ankerite							4.9		4.4
Gypsum							0.75		
Calcite							0.91		0.19
Industrial calcium oxide and carbonate							4.7	2.0	3.5
Cement									2.1
Lime kiln								0.3	0.46
Slag								2.4	
Coal fly ash								3.3	2.5
Oil fly ash								2.2	
Fused fly ash								0.3	
Glass									2.1
Miscellaneous									1.4
Petroleum catalyst cracker		8.2					0.19		3.38
Crustal, silicon-rich				11.8	10.4				
Crustal, calcium-rich				28.0	29.9				
Total crustal	17.0	24.3	31.4	39.9	40.3	42.1	11.5	28.9	30.6

Table 7. Factor loading matrices obtained by factor analysis after VARIMAX rotation

	Fine fraction				Coarse fraction			
Species	Sulfate	Steel	Auto	Unknown	Crustal-calcium	Unknown	Auto	Crustal-silicon
Mass	0.91				0.92			
C	0.68		0.65		0.88			
NO_3^-							0.87	
NH_4^+	0.94							
Al								0.95
Si	0.54	0.51						0.81
SO_4^{2-}	0.96					0.90		
Cl								
K	0.73				0.59			0.46
Ca		0.83			0.96			
Ti					0.56			0.50
Mn		0.82			0.77			
Fe		0.80			0.71			
Ni		0.93						
Cu								
Zn				0.80		0.84		
Br			0.94				0.85	
Sr					0.71	0.52		
Ba				0.93		0.96		
Pb			0.91				0.82	

a manner similar to that described elsewhere (Hopke, 1981; Currie *et al.*, 1984). The chi-square and percentage error test values suggested the presence of four detectable components in the fine fraction. Their mean concentrations and the names given them are: 8.5 μg m^{-3} 'motor vehicles,' 25.6 μg m^{-3} 'sulfate,' 1.5 μg m^{-3} 'smog,' and 3.1 μg m^{-3} 'refinery.' Their profiles are presented in Table 8. The motor vehicle component is listed in Tables 4 and 5 in the transportation category. Because the sulfate, smog, and refinery components consisted of mixtures of sulfate, nitrate, and carbon, in Tables 4 and 5 we present each component as a sum of all three species.

In the coarse fraction, five components were detected; their compositions are shown in Table 8. The concentrations of these components are shown in Tables 4–6, with the soil and limestone components combined in the general 'crustal' category. According to Table 8, the abundances of lead and bromine in the fine-fraction 'vehicles' profile and the coarse-fraction 'nitrate' profile are very similar, suggesting that both are associated with vehicles. The abundance of nitrate in the 'nitrate' profile is 458 ± 44 mg g^{-1}. If all of this were associated with $Ca(NO_3)_2$, the required abundance of calcium would be 148 ± 14 mg g^{-1}. Such a value is compatible with the calcium abundance of 88 ± 108 mg g^{-1} calculated for the 'nitrate' profile. These results are interesting, but they do not fully test the capabilities of TTFA because only 18 samples were provided. As in the case of MLR(T)/FA, at least 40 samples are needed for TTFA to be applied accurately.

Source apportionment using data from X-ray powder diffraction and microscopy

Effective variance chemical mass balance using XRD data [CMB(XRD)]. Davis of the South Dakota School of Mines and Technology (SDSM) used an effective variance CMB method similar to that used by investigators from ERT and NEA. However, the only elemental data used were the carbon concentrations supplied to participants. The other data used were XRD results for the following crystalline species: ammonium sulfate, lead ammonium sulfate, quartz, calcite, dolomite and halite. These XRD results consisted of Davis' analysis of XRD spectra provided by Thompson and Jacklevic of Lawrence Berkeley Laboratory (LBL) (see Johnson *et al.*, 1984). Davis determined the relative abundances of the crystalline species and applied a normalization procedure so that the results accounted for the measured total mass minus carbon. Profiles consisting of the abundances of seven species were constructed for the following five sources: transportation, combustion, ocean aerosol, soil and road dust. The combustion profile contained the equivalent of 12.3% lead (expressed as 12.7% lead oxide). The soil and road dust profiles were obtained by XRD analyses at SDSM of the Houston road and soil samples. The fine and coarse ambient concentrations were combined into a single fraction for which the CMB computations were performed. The combustion profile is listed in Tables 4 and 5 as the sum of the sulfate and carbon categories. Subsequent to the CMB calculations, the contributions of transportation and combustion to the fine and coarse fractions were calculated on the basis of their major components (e.g. sulfate in combustion) present in these two size fractions. Details of this procedure are reported elsewhere (Davis, 1982).

Q-mode analysis of SAX data. Johnson of the State University of New York (SUNY) applied SAX (Johnson and Twist, 1982) to characterize (under the same operating conditions) the coarse-fraction Houston ambient samples as well as a variety of source materials. The resulting 'signatures' consisted of the weight percent contribution of each of 25 particle types. A fixed classification scheme was used for all ambient and source particles, regardless of particle size. The source signatures used in the apportionment covered a variety of area and point source types. Most were obtained in and around Syracuse, NY, but the three Houston soil and road dust samples were included. Fifty-five source type signatures were available for the apportionment; using a Houston emissions inventory (Heisler *et al.*, 1981), 30 source signatures were chosen to represent the Houston area.

All of the 18 ambient and 30 source sample signatures were combined and subjected to a Q-mode factor analysis (Hall *et al.*, 1981). A VARIMAX rotaton was applied to the results, and the extremes in variable space (25 particle classes) were identified. 'Extreme' samples were those closest to the rotated factor axes and at the outer boundary of the hyperellipsoid. None of the Houston ambient samples represented extremes in variable space, which indicated that all potential source types had been captured in the data set.

Reproduced from
best available copy.

Table 8. Source profiles obtained by target transformation factor analysis

Species	Fine fraction (mg g^{-1})				Coarse fraction (mg g^{-1})				
	Vehicles	Sulfate	Smog	Refinery	Soil-silicon	Soil-calcium	Marine	Nitrate	Sulfate
C	432 ± 15	70 ± 5	323 ± 31	338 ± 27	0.2 ± 0.1	0.05 ± 0.02	0.2 ± 0.1	0.4 ± 0.1	0.1 ± 0.1
NO_3^-		8 ± 2	215 ± 13			12 ± 8		458 ± 44	140 ± 40
Na							312 ± 41	110 ± 30	44 ± 19
Al	11 ± 5				65 ± 42			98 ± 82	
Si	15 ± 2	2 ± 1	9 ± 5		140 ± 90	63 ± 29		220 ± 160	140 ± 130
SO_4^{2-}	35 ± 5	538 ± 2	42 ± 10	155 ± 9				130 ± 50	354 ± 39
Cl	3 ± 1		5 ± 2			3 ± 2	444 ± 24		
K	8 ± 2	1.2 ± 0.7			17 ± 3	2 ± 1		19 ± 6	12 ± 5
Ca	16 ± 5		33 ± 10		40 ± 55	159 ± 20		88 ± 108	95 ± 83
Ti					3 ± 1	1 ± 1			
Mn	1 ± 1				1 ± 1	0.6 ± 0.2			2 ± 1
Fe	11 ± 6				33 ± 13	17 ± 5		41 ± 24	62 ± 19
Ni	0.3 ± 0.2								
Cu	3 ± 1		5 ± 2		2 ± 1			2 ± 1	
Zn	6 ± 4		11 ± 9			2 ± 1			22 ± 3
Br	5 ± 1		5 ± 2		3 ± 1			8 ± 1	1 ± 1
Sr					0.3 ± 0.1	0.2 ± 0.1	0.6 ± 0.4	0.4 ± 0.2	1.1 ± 0.2
Ba	5 ± 2		6 ± 3					21 ± 13	37 ± 10
Pb	47 ± 10	4 ± 3			4 ± 3		1 ± 1	42 ± 6	

The 16 most extreme signatures were selected. The source apportionment consisted of a vector solution for ambient samples contained within the boundaries of the hyperellipsoid defined by these 16 extreme signatures. The normalized composition score on these 16 axes was reported as the fractional contribution of each source type to each ambient sample. The masses measured by β-gauge were then used to express the results in units of micrograms per cubic meter.

Light microscopy (LM). Crutcher of Microlab Northwest (MN) used LM to analyze one coarse-fraction ambient sample on a Nuclepore filter. Particles were first identified in 23 different classes. Then, on the basis of densities assigned to the classes and measured particle dimensions, the sample's composition was constructed. The mass concentration of each particle type was determined by multiplying the sample's total mass concentration by the particle type's fractional mass abundance.

Table 6 shows results for 10 specific crustal categories; in Table 4, these specific categories are combined into the general crustal category. The remaining 13 detected categories were condensed into general categories in the following way: the steel category was used for industrial hematite, rust, iron scale and magnetite spheres; the sulfate and carbon categories were used for soot, tire wear particles, charred wood, plant parts and pollen. Particles specific to engine exhaust were not reported.

DISCUSSION OF RESULTS

Total mass

When the Houston data set was first obtained, an attempt was made to apportion the measured mass to the measured species plus the oxygen and hydrogen assumed to relate to them (Dzubay *et al.*, 1982). It was possible to account for 78 ± 10% and 98 ± 20% of the measured fine- and coarse-fraction mass, respectively. Because the measured species plus assumed amounts of oxygen and hydrogen did not account for all of the fine-fraction mass, the way that mass was used in the model calculations affected the predictions of mass.

Three of the investigators (ERT-A, ERT-B and NEA) did not use mass as an explicit parameter in their CMB(EV) computations. (However, they did consider mass; see Step d of the CMB(EV-1) procedure described by Currie *et al.* [1984].) Their results for the fine fraction agree well with each other but only account for 70–80% of the measured fine-fraction mass, which is consistent with the fraction accounted for by the measured species.

Both MLR(T)/FA and TTFA used mass explicitly as a parameter that was fit by least-squares procedures. Consequently, the TTFA results for average mass (Table 5) are in almost exact agreement with the measured values for both the fine and coarse fractions. MLR(T)/FA permitted an intercept; its values were -0.1 and $-2.5\ \mu\mathrm{g\ m^{-3}}$ for the fine and coarse fractions, respectively. These intercepts plus the predicted component values (Table 5) agree closely with the average measured values. In CMB(XRD), mass was used implicitly in the normalization of the abundances of seven species so that their sums in the fine and coarse fractions equalled 100% of the total fine plus coarse mass. Consequently, the CMB(XRD) results agree closely with the measured fine plus coarse mass.

Transportation category

Because lead is the most important marker for the transportation category, the concentration predicted for this category is related closely to the abundance of lead in the source profile. The abundances of lead in the transportation source profiles used by the various investigators (Table 3) vary by a factor of 4, and the resulting estimates for the transportation category (Tables 4 and 5) vary by a similar factor. Such variations partly reflect the use of different definitions of the transportation category. For example, the ERT investigators used a profile for noncatalyst cars for some samples and a profile for a mixed vechicle population for others. Table 4 illustrates a case in which ERT-B used the former while ERT-A used the latter. Consequently, their results in Table 4 differ by a factor of 1.7.

The 13–14% lead content assumed for a mixed population of vehicles and used in some of the CMB(EV) computations differs greatly from the 4–5% deduced by MLR(T)/FA and TTFA. The former is based on source measurements made before 1979 and may overestimate the lead content during 1980. Data on lead consumption by vehicles indicate a 50% decrease between 1978 and 1980 (Du Pont, 1982). The MLR(T)/FA and TTFA profiles are based on ambient measurements and may better reflect the actual lead content of the vehicle exhaust. Also, MLR(T)/FA and TTFA may include secondary aerosol associated with vehicles; such secondary aerosol cannot conveniently be included in the profiles used in CMB(EV) analyses. Road and tire wear products plus vehicular entrained soil could also be included in the MLR(T)/FA and TTFA profiles, but such substances are emitted primarily in the coarse fraction. From the TTFA results shown in Table 8, one can infer that the fine-fraction 'vehicles' component contains less than 15% crustal road dust. If the transportation profile deduced by TTFA is correct, then substantial increases in the CMB(EV) estimates for the contributions of vehicles to ambient air would be necessary. Overall, the MLR(T)/FA and TTFA results for the transportation component are about a factor of 3 higher than those deduced by CMB(EV).

The CMB(XRD) results for the transportation category are based upon the concentrations of lead ammonium sulfate, the marker species that was used. Because lead is emitted from vehicles as halides, chemical reaction with atmospheric acid sulfate particles is required to produce lead ammonium sulfate. Thus, the CMB(XRD) results depend upon the completeness of such chemical reactions, which is affected by factors not associated with vehicular sources. Variability in the completeness of the chemical reactions would cause variability in the CMB(XRD) results for the transportation category. Perhaps the results would be improved if the lead concentrations obtained by XRF were used in the analysis.

Sulfate and carbon categories

Because sulfate, nitrate and carbon all occurred in three separate categories deduced by TTFA in the fine fracton (sulfate, smog and refinery), it was not possible to resolve them into the chemically pure categories of Tables 4 and 5. Thus, results may be compared only for the sum of the carbon, sulfate and nitrate categories. On that basis, the agreement is generally very good, with the MLR(T)/FA and TTFA results 10–20% higher than the CMB(EV) results based upon elemental compositional data. Similarly, the CMB(XRD) results are higher than the CMB(EV) results owing to the inclusion of a data normalization step in the CMB(XRD) procedure.

Crustal component

The crustal component, which constitutes at least 70% of the coarse-fraction mass concentration, is shown in detail in Table 6. The XRD measurements by LBL are in fairly good agreement with the LM data for quartz and calcite, but XRD did not detect the large clay component reported by LM. The numbers of crustal components determined by the various methods are: 10 by LM, 8 by SAX, 5 by XRD, and 2 by CMB(EV), MLR(T)/FA, TTFA and CMB(XRD). It is encouraging to note that SAX and LM together resolved $\sim 12\ \mu\text{g m}^{-3}$ or 40% of the crustal component into the following specific source classes: cement, slag, fly ash, industrial calcium oxide and carbonate, lime kiln and glass. There was good agreement between LM and SAX for the first three source classes but not for the last three. Only SAX could resolve fly ash into components from coal and oil. Several of the additional components reported by XRD, SAX and LM (such as clay, quartz and calcite) may be constituents of the two crustal components reported by CMB(EV), MLR(T)/FA, TTFA and CMB(XRD). Further investigation is needed to confirm the crustal components.

'Soil, silicon-rich' and 'soil, calcium-rich' are the two distinct crustal components detected by TTFA; their compositions are similar to those of the soil and road dust samples collected in Houston. However, we cannot draw specific conclusions from this similarity because it is not known how accurately the three soil and road samples represent the thousands of soil sites and roads than can contribute to Houston's aerosol.

Composite receptor model results

'Composite receptor modeling' refers to the application of two or more receptor models to the same set of aerosol samples. The Quail Roost II exercises, obviously, afford an opportunity to compile and evaluate composite results for Houston aerosol. Table 9 shows results in 14 categories obtained by combining TTFA, LM, SAX and emissions data. In the fine fraction, the TTFA result for vehicle exhaust is resolved into diesel and gasoline powered vehicle components. This resolution is based on a 6:1 ratio between emission rates for gasoline and diesel powered vehicles in the Houston airshed (NEDS) and on the assumption that this airshed average accurately represents the impacts of both vehicle types at our sampling site. The component labeled 'other fine' represents the sum of the 'refinery' and 'smog' components identified by TTFA.

In the coarse fraction, nine components pertaining to specific emission source types were determined by SAX and LM. The component labelled 'other coarse' consists mainly of crustal matter and represents the difference between the measured mass concentration and the sum over the other components in the coarse fraction. Although SAX, LM and XRD identified several specific components such as limestone, quartz, clay, and gypsum, these are included in 'other coarse' because they are not specific to a particular emission source

Table 9. Composite receptor modeling results for the 12-h period beginning at 0700 on 10 September 1980

Size range	Component	Concentration* (μg m^{-3})	Method
Fine	Vehicle exhaust, diesel	2.5	TTFA and emissions ratio
	Vehicle exhaust, gasoline	15.1	TTFA and emissions ratio
	Sulfate and related species	35.2	TTFA
	Other fine	1.1	TTFA
Coarse	Marine	0.3	TTFA
	Sulfate and related species	2.5	TTFA
	Nitrate and related species	0.6	TTFA
	Pollens and spores	1.0	LM
	Other biological matter	0.3	LM
	Tire wear and asphalt	1.6	LM
	Lime kiln	2.4	SAX
	Slag	2.9 (0.4)	LM, SAX
	Coal fly ash	2.2	SAX
	Oil fly ash	0.3	SAX
	Cement	0.4 (0.1)	LM, SAX
	Steel, rust	3.1 (0.1)	LM, SAX
	Other coarse	17.2	

* Each result based on both LM and SAX is an average, and each value in parentheses is half the difference between the SAX and LM values.

type and could have been emitted from a diverse assortment of sources (such as wind erosion of soil, suspension of road dust, and manufacturing processes).

CONCLUSIONS

Fairly good agreement is evident when the source apportionment results are compared in eight general categories. Differences among results by the various methods are often attributable to differences in the source signatures that were selected. The participants' results for major categories such as sulfate plus carbon and crustal matter agree within a range of 30–40%. However, the results for the transportation category obtained by multivariate analysis are about a factor of 3 greater than the results obtained by CMB for a mixed population of vehicles. The latter were determined mainly by the abundance of lead in the source signature for vehicle exhaust. The signature used was obtained about 2 y before the Houston study and does not take into account the factor-of-2 decrease in the lead content of vehicle exhaust that occurred between 1978 and 1980. In addition, the results by multivariate analysis may include secondary vehicle aerosol, which was not included in the signatures used by CMB. TTFA results indicated that the transportation category in the fine fraction contained less than 15% crustal matter. Transportation source signatures derived by TTFA are potentially more realistic than those for weighted populations of vehicles derived by means of emissions inventories.

Although apportioning aerosols into eight categories is of great value, obtaining categories that are more specific to known emission source types would be desirable. In fact, LM and SAX did successfully resolve the general categories into several specific source related categories. TTFA, SAX, and LM (if their results are considered in combination) resolved the aerosol mass concentration into about 14 categories. Had source signatures specific to Houston been available, it is likely that the aerosol could have been resolved into even more components.

Although the present study provides valuable insight into the relative performances of several receptor models, certain limitations in the study prevented a demonstration of the models' full capabilities. One limitation was the lack of source profiles specific to Houston. Most of the profiles were obtained in other cities, making their applicability to Houston only an assumption. Three soil and road dust profiles for Houston were provided, but whether these profiles were truly representative of the soil and road dust suspended in the Houston atmosphere was not known. A second limitation was the short time between the receipt of data and the date for reporting results, which made it difficult for investigators to devote enough time to this project. A third limitation was that the 18 sampling periods were too few for MLR(T)FA and TTFA. These methods require at least 40. A fourth limitation was that the samples had mass loadings that were too low for XRD to demonstrate its full analytical capability. Overall, this intercomparison was valuable because it was the first of its kind and provided impetus to better define the data requirements of receptor models.

A more comprehensive intercomparison of receptor models would be useful. Such an intercomparison could involve receptor models that require more comprehensive data than were available for Houston. The following should be considered in the design of such an intercomparison:

(1) Participants should be involved in the design and planning.

(2) The data on which the receptor models are to be applied should be obtained in a manner consistent with the guidelines of Gordon *et al.* (1984).

(3) The ambient aerosol data set should include at least 40 sampling periods.

(4) All participants should receive the complete data set, including XRD, SAX and LM data on both source and ambient samples.

(5) Participants should receive all data at least nine months before results are due in order to insure ample time to develop models that include all parameters measured.

(6) A protocol specifying how mixed vehicle populations are to be reported should be developed and provided to participants.

(7) The representativeness of any soil and road dust profiles provided should be ascertained.

REFERENCES

Cass G. R. and McRae G. J. (1981) Source–receptor reconciliation of South Coast Air Basin particulate air quality data. Report for California Air Resources Board.

Cooper J. A., Watson J. G. and Huntzicker J. J. (1979) Summary of the Portland Aerosol Characterization Study. In *Proc. Air Pollution Control Association*, 79–24.4. Air Pollution Control Association, Cincinnati, Ohio.

Currie L. A., Gerlach R. W., Lewis C. W., Balfour W. D., Cooper J. A., Dattner S. L., De Cesar R. T., Gordon G. E., Heisler S. L., Hopke P. K., Shah J. J., Thursten G. D. and Williamson H. J. (1984) Interlaboratory comparison of source apportionment procedures: Results for simulated data sets. *Atmospheric Environment* **18**, 1517–1537.

Davis B. L. (1982) Hybrid Model for Source Apportionment of the Houston Aerosol for the Period 10–18 September 1980. Report 82-3, Institute of Atmospheric Sciences, South Dakota School of Mines and Technology, Rapid City, South Dakota.

E. I. du Pont de Nemours & Co., Inc. (1982) Statement presented to U.S. Environmental Protection Agency at a Hearing on Regulation of Fuel and Fuel Additives, Lead Phasedown Regulation, Notice of Proposed Rulemaking, Washington, D.C., 15–16 April, Docket Number A-81-36.

Dzubay T. G., Stevens R. K., Lewis C. W., Hern D., Courtney W. J., Tesch J. W. and Mason M. A. (1982) Visibility and aerosol composition in Houston, Texas. *Envir. Sci. Technol.* **16**, 514–525.

Full W. E., Ehrlich R. and Klovan J. E. (1981) Extended Q model—Objective definition of external end members in the analysis of mixtures. *Math. Geol.* **13**, 331–344.

Gordon G. E., Pierson W. R., Daisey J. M., Lioy P. J., Cooper

Reproduced from best available copy.

J. A., Watson J. G., Jr. and Cass G. R. (1984) Considerations for design of source apportionment studies. *Atmospheric Environment* **18**, 1567–1582.

Hare C. T. and Black F. M. (1981) Motor vehicle particulate emission factors. Papaer 81-56.5, 74th Annual Meeting and Exposition of the Air Pollution Control Association, Philadelphia, Pennsylvania.

Heisler S. L., Watson J. G., Shah J. J., Chow J. C., Collins J. C. and Whitney J. (1981) Recommendations for the Design of Aerosol Characterization Studies in Houston and El Paso, Texas. Document P-A716-110, prepared for Texas Air Control Board, Austin, Texas. Environmental Research and Technology, Concord, Massachusetts.

Henry R. C., Lewis C. W., Hopke P. K. and Willliamson H. J. (1984) Review of receptor model fundamentals. *Atmospheric Environment* **18**, 1507–1515.

Hopke P. K. (1981) The application of factor analysis to urban aerosol source resolution. In *Atmospheric Aerosol: Source/Air Quality Relationships* (edited by Macias and Hopke), pp. 21–49. ACS Symp. Series, No. 167. American Chemical Society, Washington, D.C.

Johnson D. L. and Twist J. P. (1982) Statistical considerations in the employment of SAX results for receptor models. In *Receptor Models Applied to Contemporary Pollution Problems, Proc. APCA Speciality Conf.* SP-48 (Danvers, Massachusetts, October 1982), pp. 224–237. Air Pollution Control Association.

Johnson D. L., Davis B. L., Dzubay T. G., Hasan H., Crutcher E. R., Courtney W. J., Jaklevic J. M., Thompson A. C. and Hopke P. K. (1984) Chemical and physical analyses of Houston aerosol for interlaboratory comparison of source apportionment procedures. *Atmospheric Environment* **18**, 1539–1553.

Kleinman M. T., Pasternack B. S., Eisenbud M. and Kneip T. J. (1980) Identifying and estimating the relative importance of sources of airborne particulates. *Envir. Sci. Technol.* **14**, 62–65.

Lang J. M., Snow L., Carlson R., Black F., Zweidinger R. and Tejada S. (1981) Characterization of particulate emissions from in-use gasoline-fueled motor vehicles. Paper 811186, SAE Technical Series, Society of Automotive Engineers, Fuels and Lubricants Meeting, Tulsa, Oklahoma.

Pierson W. R. (1979) Particulate organic matter and total carbon from vehicles on the road. In *Proc. Carbonaceous Particles in the Atmosphere* (edited by Novakov), pp. 221–228. LBL-9037, CONF-7803101, UC-11. Lawrence Berkeley Laboratory, Berkeley, California.

Shah J. J. (1981) Measurements of carbonaceous aerosol across the U.S.: Sources and role in visibility degradation. Ph.D. dissertation, Oregon Graduate Center, Beaverton, Oregon.

Stevens R. K. and Pace T. G. (1984) Overview of the Mathematical and Empirical Receptor Models Workshop (Quail Roost II). *Atmospheric Environment* **18**, 1499–1506.

Watson J. G. (1979) Chemical element balance receptor model methodology for assessing the sources of fine and total suspended particulate matter in Portland, Oregon. Ph.D. dissertation, Oregon Graduate Center, Beaverton, Oregon.

Watson J. G. *et al.* (1981) Receptor models relating ambient suspended particulate matter to sources. EPA-600/2-81-039, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Williamson H. J., Balfour W. D. and Schmidt C. E. (1982) Source apportionment through weighted-ridge regression. Unpublished report, Radian Corporation, Austin, Texas.

Atmospheric Environment Vol. 18, No. 8, pp. 1567–1582, 1984
Printed in Great Britain.

CONSIDERATIONS FOR DESIGN OF SOURCE APPORTIONMENT STUDIES[*]

GLEN E. GORDON[†]
Department of Chemistry, University of Maryland, College Park, MD 20742, U.S.A.

WILLIAM R. PIERSON
Engineering and Research Staff, Ford Motor Company, Dearborn, MI 48121, U.S.A.

JOAN M. DAISEY, PAUL J. LIOY
Institute of Environmental Medicine, New York University Medical Center, New York, NY 10016, U.S.A.

JOHN A. COOPER
NEA, Inc., Beaverton, OR 97005, U.S.A.

JOHN G. WATSON, JR.
Desert Research Institute, Reno, NV 97005, U.S.A.

and

GLEN R. CASS
Department of Environmental Engineering Science, California Institute of Technology, Pasadena, CA 91125, U.S.A.

(*First received* 9 *June* 1983 *and in final form* 19 *January* 1984)

Abstract—This report recommends procedures for source and ambient sampling and analysis in source apportionment studies. The recommendations are based on the results of receptor model studies of atmospheric particles in urban areas, especially a recent study of Houston, TX, undertaken as part of the Mathematical and Empirical Receptor Models Workshop (Quail Roost II). The recommendations are presented at three levels of increasing cost and detail of information obtained. Existing mass emissions inventories combined with chemically resolved test data from similar sources (not necessarily in the same locale) can be used to initially estimate the sources of elements present on ambient particles. To aid local users in construction of chemically resolved emission estimates, the U.S. Environmental Protection Agency (EPA) is compiling a library of compositions and size distributions of particulate emissions from major source types. More reliable source characterization can be achieved if the actual sources are tested directly. EPA should develop and publish detailed procedures for source sampling that would be more appropriate for receptor model use than are existing standard methods. Source and ambient sampling should be conducted by similar methods. If possible, particles from sources should be collected in a way that simulates changes that would normally occur before they reach distant receptors (e.g. by diluting and cooling the particles from hot sources). It is recommended that particulate samples be routinely collected in two size fractions by use of virtual impactors, and that all samples be subjected, at a minimum, to mass and X-ray fluorescence analyses. Additional measurements are suggested for obtaining more detailed information: neutron activation analysis; X-ray diffraction; automated particle classification by electron microscopy; analyses for classes of organic species, ^{14}C and thermally released carbonaceous species; and real-time observation of several gases during sample collection. Methods for collecting meteorological data in parallel with ambient samples are described, as are methods for incorporating such data into the source identification process.

Key word index: Ambient sampling, receptor modeling, sampling strategies, source apportionment, source characterization, source emissions, source sampling.

INTRODUCTION

This paper is the last in a series on the use of multispecies analyses of urban atmospheric particulate material and treatment of the results by receptor models in order to identify sources of the particles and estimate their contributions to ambient particulate levels. The entire series resulted from discussions and comparisons of analytical and interpretive methods conducted at the Mathematical and Empirical Receptor Models Workshop (Quail Roost II) (Stevens and Pace, 1984). In part, the Quail Roost II exercises involved application of various receptor models to ambient aerosol data obtained in Houston, TX. Johnson *et al.* (1984) describe the collection and analysis of particles from the Houston atmosphere and Dzubay *et al.* (1984) examine the receptor model results obtained for that data set.

---

[*] A Report from the Mathematical and Empirical Receptor Models Workshop (Quail Roost II).

Author to whom requests for reprints of this article should be addressed. Requests for reprints of the entire series of six articles (including this one) documenting the Quail Roost II workshop should be directed to Robert K. Stevens, MD-47, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, U.S.A.

This paper sets forth recommendations for the design of receptor model studies of other urban areas. These recommendations are based on the results of the Quail Roost II study and of previous studies in other urban areas. We discuss improvements needed for future tests of receptor models and for application of receptor models to practical problems. As resources needed for an ideal study are rarely available, we suggest various types of source and ambient measurements that are appropriate at different resource levels.

The types of source measurements discussed include direct measurements at the source in question, the resolution of source terms from ambient sampling data by target transformation factor analysis, and a wind trajectory analysis method for identifying samples from ambient networks that are heavily influenced by specific sources. The discussion of ambient sampling covers not only the types of filters, samplers, etc. to be used, but also the need for adequate meteorological data and the methods needed for individual compounds in the particulate and gas phases.

SOURCE CHARACTERIZATION

Direct measurements

To identify and catalog source signatures, one must consider the following parameters: the number of sources present; the time (calendar dates) when tests and analyses were performed; the presence of unique pollutant emitters; the variability in emissions composition of each source type; and the similarity of compositions of emissions from different types of sources. From these and other concerns peculiar to each site, a hierarchy of decisions must be made based on the questions to be answered and the resources available for the study. For example, if only compliance with a mass concentration standard must be determined, only the source types that contribute the majority of ambient mass may need to be assayed. However, if contributions to the concentration of a specific chemical species are needed, more detailed source signatures are required.

Resources are an important consideration in developing a receptor modeling program for any airshed. The following outline of three resource levels can be used to achieve various degrees of sophistication in a receptor model application. Each level depends upon completion of the previous level. Each level requires more resources than the preceding one, but provides receptor model results of greater reliability. An initial application of the receptor model using source composition data in each level may direct the actions and minimize the costs in subsequent levels. Obviously, we cannot give a single protocol that will be applicable to all locations and meteorological conditions, so the investigator must use judgment in deciding if some changes are needed for a particular study.

Level I. Minimal resources. The first step towards source characterization is to complete a list of the point and area sources most likely to affect pollutant concentrations at the sampling sites. Properties of their emissions can be estimated from available emissions inventories, site surveys, local knowledge and the results of previous tests of similar types of sources, not necessarily in the same locale.

For point sources, the National Emissions Data System (NEDS) and the local agency emissions inventories, from which NEDS is derived, are the best starting points. For each source emitting $> 1\ \mathrm{t\,y^{-1}}$ of SO_2, NO_x, CO or total emitted particulate matter (TEP), the NEDS condensed point source inventories contain source classification codes (SCCs) that identify the fuels used and processes involved, universal transverse Mercator (UTM) coordinates of the point of emissions, and calculated and allowable emission rates of SO_2, NO_x, CO, TEP and hydrocarbons. These records are obtainable in computer-compatible formats that can be sorted by SCC and, subsequently, by TEP emission rates within each SCC. Important point sources of a given type should be identified on the basis of SCC and estimated TEP emission rate. The lower cutoff of emission rates will depend on the resources available. Past experience shows that most inventoried sources with TEP rates of $< 50\ \mathrm{t\,y^{-1}}$ can be ignored unless they release fugitive emissions with low plume rise. The TEP rate is often inaccurate and should be treated as an order-of-magnitude estimate.

Individual sources and receptor locations should be plotted by UTM coordinates on a map of the area. If the intent is to determine contributions from individual sources, this map can be used to place receptors strategically. With knowledge of prevailing wind directions, one can place receptors so as to maximize the probability of collecting particles from the sources of major interest (see Wind Trajectory Analysis).

Area sources most likely to contribute to a receptor are those in its immediate environs. These sources are best identified from a survey of the area around the site. This survey, based on the National Air Monitoring System hardcopy site survey, should include:

(1) A street-and-block map for the 2-km radius around the sampling site that shows typical traffic counts, curbing, lights and general dirtiness of roadway sections; block classifications by residential, commercial, industrial, and agricultural categories; and locations and dimensions of vacant lots (with coverings), storage piles and parking lots.

(2) A description of the sampling station, including

the sampler layout; elevation of sampling probes above sea level, ground level and platform level; UTM coordinates; and a summary of nearby structures that might affect micrometeorology.

(3) Photographs in eight cardinal directions from the sampling platform.

The point source summaries and site surveys should be reviewed by a knowledgeable person from the area. Local air pollution specialists are aware that emission rates often change more rapidly than indicated in emissions inventories and dated site surveys.

Source compositions for chemical components and particle size fractions under conditions compatible with those prevailing at the receptor should be compiled from existing data for the source types in the area. This compilation would be assisted by a documented source characterization library. In such a library, data for each source should include its type, specific identifier, and the time, place and circumstances of the source test; the characteristics of source operation and control devices used during the test; the sampling methods used for each test; the chemical species and size ranges measured and the analytical methods used and concentrations and uncertainties therein for designated chemical species by size range. Some of this information is compiled in the Fine Particle Emissions Information System (FPEIS) (Schrag and Johnson, 1978).

Within the U.S. Environmental Protection Agency (EPA), the Air Monitoring Techniques Branch of the Office of Air Quality Planning and Standards recently initiated compilation of a source composition library that should contain sufficient information to allow selection of the source signatures that are most similar to those in the local area. At present, no existing document or computer library contains all available data on source signatures; consequently, data must be assembled from the relevant literature. Such literature includes previous receptor model studies in which the components are listed (Kowalczyk *et al.*, 1982; Watson, 1979; Cass and McRae, 1981; Alpert and Hopke, 1980, 1981), published articles of reports of source studies, and soil analyses from the area based on national compilations (Shacklette *et al.*, 1971) or those of state geological surveys. As discussed below, however, the airborne soil probably has a composition different from that of bulk soil. It may be preferable to construct an approximate component for emissions from coal-fired plants rather than to determine the composition directly from studies on plants from another area. Such a component could be compiled on the basis of the composition of coal consumed in the local area as modified by enrichment factors determined for elements between coal and emitted particles (Gordon *et al.*, 1981).

Most local areas will have insufficient funds for detailed studies of motor vehicle emissions, but a reasonable estimate of their compositions can be made on the basis of studies by Pierson and Brachaczek (1983), Ondov *et al.* (1982), and older work reported in the literature. As the compositions and masses of particles from diesel engines are quite different from those of gasoline-burning engines, the amounts of the different types of vehicle traffic in the area should be estimated from traffic counts or fuel sales.

Initial receptor modeling using these Level I source characteristics and receptor measurements can guide the efforts of Level II source characterization by evaluation of their adequacy for the models being applied.

Level II. Moderate resources. The accuracy and precision of estimates of source contributions can be improved if actual compositions of local source emissions are determined (instead of using measurements of similar sources in other locations). This section outlines a general method to be followed.

First, identify source categories and specific emitters that previous modeling or past experience have shown to be potential contributors. Major sources of mass include wind-raised or vehicle-entrained dust from agricultural soils, roadways and parking lots; active storage piles (i.e. piles where materials are frequently added or removed); transfer processes, such as loading and unloading of vehicles; and sources such as power plants, smelters, steel plants, cement plants and refuse incinerators.

Second, from each selected source obtain 'grab' samples that will best represent the average composition. Here 'grab' sampling means the collection of bulk material from a pollution control device or from piles of material that may produce fugitive emissions, or the simple in-stack collection of suspended particles. Sampling methods used at each source should consider temporal variations caused by the process schedule and fuel switching.

Third, try to isolate the portion of the bulk sample that is likely to remain airborne long enough to reach a receptor. Ordinarily, it is inappropriate to analyze a bulk sample. A convenient and necessary (but not always sufficient) procedure is to sample in a suspension chamber after drying the bulk sample and sieving it through a standard Tyler 400-mesh (36-μm) screen. The resuspended sample should be collected on a filter of high purity (cellulose, polycarbonate or Teflon) with a loading appropriate for the material to be measured and the intended analytical method. For example, a quartz filter cannot be used if silicon is to be measured.

If possible, the same size intervals used for ambient sampling should be used for source sampling. If sources are to be controlled to meet federal standards, the size intervals may be established by those standards. If not, the best scientific judgement should be followed, using available and convenient sampling methods. One useful approach is to employ virtual impactors (dichotomous samplers) to collect source and ambient samples in two size intervals (< 2.5 μm and 2.5–10 μm).

For grab sampling of in-stack suspended particles, a variation of EPA Method 5 (Rom, 1972) is used in which the filter of Method 5 is replaced by a particle-

Reproduced from
best available copy.

sizing device, such as a cascade impactor (Pilat *et al.*, 1970) or in-stack virtual impactor (Houck *et al.*, 1982). In classical source emissions tests, a pitot tube must be used to determine the velocity of stack gases at many points along two perpendicular diameters across the stack. Then the mass loading must be measured at several points. However, for receptor modeling, collecting samples at one point somewhat away from the inner wall is usually sufficient. A pitot tube is used to determine stack gas velocity at that point, and the pumping speed through the sampler is adjusted to ensure isokinetic sampling. Several examples of stack studies are described by Klein *et al.* (1975), Gladney *et al.* (1976), Greenberg *et al.* (1978a,b) and Ondov *et al.* (1979a,b).

In the discussion of Level III procedures below, we recommend that stacks be sampled by dilution methods. If one owns the equipment needed for dilution sampling, the incremental cost of dilution sampling versus the methods described above will usually be a small fraction of the cost of stack sampling; in such cases, dilution sampling is recommended.

Finally, the samples should be analyzed for species of interest. At this second level of resources, consideration should be given to additional analyses that could improve the overall results, even though these additional species have not yet been measured at the receptors. Analyses should be considered that would provide data compatible with ambient data on the following additional source characteristics: (1) total mass; (2) total mass within the 'fine' and 'coarse' size fractions; (3) individual elements and ions, especially major contributors to total mass or tracers for major sources; (4) inorganic and organic compounds, again emphasizing tracers for major sources; (5) visual characteristics, using microscopy of various types and (6) crystal structures as observed by X-ray diffraction (XRD). (Note that we discuss analytical procedures below in connection with ambient samples. As virtually unlimited quantities of grab sample are usually available, one can assume that any analytical procedure appropriate for ambient samples can also be used here.)

This second level provides source samples similar to those taken at receptors at approximately the same cost per sample. Sources of some types may not be amenable to grab sampling, or the grab sample may not be representative of the emissions that reach the receptor. More complex and costly procedures are required to characterize these sources. Sampling only in stacks and pollution control devices will miss fugitive emissions, which often contribute more suspended particulate matter, especially in the coarse fraction, than do ducted emissions.

Level III. Maximum resources. Level III testing should be confined to those sources for which it is absolutely necessary, such as major contributors that are not amenable to grab sampling. In all cases, time variations in composition and mass should be determined for each source studied. The three types of Level III tests are:

(1) Ground-based sampling of plumes. This sampling should be completed for a source whose ground-level contribution is greatest. This type of test is applicable to sources such as motor vehicle emissions and field burns. The collection devices should be the same as those used at receptor sites. A variation of this method applicable to point sources (wind trajectory analysis) is discussed below.

(2) Dilution stack sampling. This technique should be used for vented emission sources and have appropriate constraints for assuring the collection of condensable species (Heinsohn *et al.*, 1980; Houck *et al.*, 1982). Samples should be collected in size fractions similar to those collected at the receptor sites. Presently, 2.5- and 10-μm dia. size cuts appear appropriate for the anticipated particulate matter standard. In the case of high-temperature stacks, provision should be made for diluting stack gases to ambient temperature and relative humidity to allow condensation of moderately volatile species. Standardization of the dilution sampler also is necessary, especially to determine the sample size required for the receptor application. If dilution and cooling are not possible, it is desirable to collect moderately volatile species (e.g. Hg, As, Se, halogens, organics) in appropriate columns or traps downstream from filters (e.g. activated charcoal) in order to determine an upper limit on the amount that might condense in the plume at ambient temperatures (Germani, 1980). As these methods are still under study and improvement, no 'standard' procedure can be described yet.

(3) Motor vehicle emissions. Ideally, one should be able to determine separate contributions from three vehicle-related components: emissions from diesel engines, from leaded gasoline engines, and from nonleaded gasoline engines. However, except for Pb and Br from leaded gasoline, these components are not well distinguished in an urban mix on the basis of elements *per se*. As other particle characterization methods (e.g. organic analyses, microscopy, etc.) are improved, they should be used to determine better markers for these components. In such studies, it is essential that hot exhaust emissions be cooled and diluted as discussed above. Alternatively, total emissions from vehicles can be measured effectively by studies in traffic tunnels, but the data must be collected and interpreted carefully to eliminate traffic-entrained soil from the true vehicle components (Ondor *et al.*, 1982). If there is sufficient variation of the diesel/gasoline traffic, one can probably resolve these components (Pierson and Brachaczek, 1983); however, it is difficult to see how one could separate leaded and nonleaded gasoline emissions in this way. Even if one could get an accurate count of these vehicle classes, there would probably be too little hour-to-hour variation of the ratio to permit resolution of their emissions.

(4) Airborne sampling of plumes. Aircraft sampling of plumes from major sources can often be done quite

effectively. If a fixed-wing plane is used, it should have as slow a stall speed as possible to maximize time spent in plumes and to decrease corrections needed for the response times of instruments that detect the plume. The aircraft should have two or four engines to leave the area around the fuselage free of propeller wash and engine exhaust. Collection probes should be designed to ensure that samples are collected isokinetically at a specified air speed, and that the intake is in a region of laminar flow, unperturbed by the plane. It should be possible to shut off filters (e.g. when outside of the plume) and to change filters during the flight. Sampling aircraft are usually equipped with real-time detectors of species indicative of the plume (e.g. SO_2, NO_x, condensation nuclei), and data-loggers and/or strip charts so as to relate amounts of particulate matter collected to other plume indicators. Response times of real-time instruments should be as short as feasible (1-10 s), to give sharp changes at plume boundaries. It is expensive to modify a plane for sampling (usually requiring additional electric power and vacuum), and difficult to obtain FAA approval of the changes. However, once the changes are completed, plume sampling can be done very efficiently: under favorable conditions, plumes from several plants together with appropriate backgrounds can be sampled per day. The samples obtained are more appropriate for receptor model use than are stack samples, and no cooperation of personnel at the source is required. A major limitation of aircraft sampling has been that no sampling device can collect material isokinetically in a size-segregated mode at a sufficiently high rate to provide enough material for many of the analytical techniques after just a few plume traverses. Perhaps the high-volume virtual impactor being developed and tested by Solomon *et al.* (1983) will be useful in this application. Another problem, especially for sources in the midst of heavily populated areas, is that FAA restrictions may make it impossible to sample the plume at optimal locations or altitudes. An example of aircraft plume sampling of several copper smelters is the work of Small *et al.* (1981). Tethered balloon sampling systems offer an alternative method for sampling plumes (Armstrong *et al.*, 1981). Such systems allow rapid horizontal and vertical mobility for plume sampling and can be used to collect samples over longer time periods than are practical with planes.

Procedures required. If receptor models are to be widely used by state and local air pollution authorities, EPA needs to develop standard procedures for obtaining the data that are used in these models. In some cases, especially for Level III procedures, research will be needed before the procedures can be written.

Level I procedures should cover the following areas: (1) reducing NEDS Point Source Inventories for receptor model applications; (2) making surveys of areas around receptor model sampling sites; (3) using the Receptor Model Source-Composition Library and (4) assembling source compositions from the Library that are appropriate to the local area. Level II procedures should provide guidance for: (1) obtaining representative grab samples from storage piles, baghouses, roadways, parking lots, and exposed land; (2) drying, sieving and resuspending grab samples for specific analyses and (3) taking size-segregated stack samples. Level III procedures should cover these areas: (1) ground-based sampling for source emissions characterization; (2) dilution sampling of stacks and ducts for source emissions characterization and (3) airborne sampling of elevated plumes for emissions characterization.

Estimation of source signatures by multivariate methods

For the Quail Roost II workshop, a multivariate method, target transformation factor analysis (TTFA) (see Henry *et al.*, 1984), was used to estimate source profiles from the ambient data. Once source profiles have been determined, ordinary chemical mass balance (CMB) methods may be used to determine source impacts. This procedure is valuable if the only available source profiles are from source tests conducted in other geographic areas, or if it is reasonably certain that the source profile has changed or is highly variable. TTFA can be used to adjust an assumed source profile to maximize that profile's consistency with observed intercorrelations. Thus, TTFA can improve the source profile before ordinary CMB analysis is attempted. If sufficient data are available (30 or more degrees of freedom per variable), TTFA can provide a useful check on assumed source compositions. If no source composition data are available, TTFA can provide quantitative estimates of source profiles. For instance, if a previously unknown source is deduced, its approximate composition can be determined.

Certain dangers are associated with the use of TTFA to obtain source signatures, especially the 'smearing' of signatures between sources whose emissions are highly correlated in time or space. Hopke and Gordon (unpublished results) have compared results from CMB (Kowalczyk *et al.*, 1982) and TTFA for 130 samples from the Washington, DC area. Signatures obtained by TTFA for marine aerosol and incinerator emissions contained admixtures of the signature of oil-fired power plants (indicated by high V and Ni concentrations). This smearing of signatures, which resulted from correlations among those source emissions, reflects the close proximity of an oil-fired plant to an incinerator in Washington and the fact that large oil-fired plants are located to the south and east of Washington, so that air parcels coming from those directions, bearing marine aerosol, also contain particles from oil combustion. A comparison of signatures of sources in St. Louis obtained by TTFA (Alpert and Hopke, 1981) and wind trajectory analysis indicates a similar smearing of signatures by TTFA among sources that are located in the same general direction with respect to the receptor. In cases of large ambient data sets that include extensive meteorological data (especially wind direction), the smearing problem may be avoided by selectively applying TTFA to only the samples collected under specified wind conditions.

Wind trajectory analysis

Gordon (1980) has reported on the use of wind trajectory analysis (WTA) for obtaining compositions of particles from specific sources by using large ambient data sets that contain extensive wind direction date—in particular, the St. Louis Regional Air Pollution Study (RAPS) data set (Stevens *et al.*, 1978; Loo *et al.*, 1978; Dzubay, 1980). The method has been further developed by Rheingrover and Gordon.

In WTA, the following steps are performed for each species in each size fraction:

(a) Determine average concentration and standard deviation at each station.

(b) Search the data from each station for sampling periods that meet two conditions: a concentration of $\sim 3\sigma$ above average for the station; and fluctuations of surface wind direction, σ_θ, of less than $\sim 20°$.

(c) For sampling periods that meet these criteria, construct a histogram of mean wind direction. If there are dominant sources of the species, the histogram will display clusters about angles pointing toward the sources.

(d) Using the mean angles from the clusters, construct a map of back trajectories observed at all stations in the network. A convergence of trajectories from two or more stations often indicates the presence of a dominant source of the species near the convergence.

(e) For samples identified as heavily influenced by a particular source, construct linear regressions of concentrations of all measured species vs those of one or more species strongly associated with particles from the source. The correlation coefficients indicate the likelihood that the various species are present on particles from the source, and the regression coefficients yield relative concentrations. If mass contributions from the source are large enough to give good correlations with the species in question, absolute concentrations of those species in the particulate matter can be obtained.

Rheingrover and Gordon (1980) used WTA to identify samples influenced by the following point sources in St. Louis: an iron and steel complex, an iron works, a zinc smelter, two lead smelters, two copper plants, an incinerator, a pigment plant, and a fertilizer plant. Surprisingly, they also found trajectories (apparently to areas of high traffic density) that were associated with motor vehicles. For most sources, concentrations of 6–10 elements relative to that of a prominent element could be determined.

WTA has several advantages relative to traditional methods:

(1) It avoids the cost of source measurements.

(2) The source signatures obtained are characteristic of particles *as received* at the receptors. For example, volatile elements will have condensed (if they are going to), and very large particles will have fallen out.

(3) Problems that could arise by use of different sampling equipment or analytical methods at the source vs the receptors, or by use of measurements widely separated in time, are avoided.

(4) Fugitive emissions, which are difficult to measure by conventional methods, can often be observed.

(5) Cooperation of personnel at the source is not required.

The use of linear regressions to obtain the compositions of source particles presupposes that fluctuations about the regression lines are random. However, aside from experimental error, the fluctuations result from real variations of intensities of components from other sources. Consequently, one should be able to resolve the other components by a multivariate method to leave a more accurately known component for the source in question. Chang *et al.* (1982) have subjected the groups of St. Louis trajectory samples identified by WTA to TTFA to obtain more comprehensive, reliable composition patterns for the sources identified.

Another feature of WTA is just beginning to be exploited. X-ray fluorescence (XRF) is a rapid, inexpensive method for analyzing vast numbers of samples that can be used to provide data for WTA. The latter can identify the samples containing large contributions from specific sources. The samples so identified can be subjected to more elaborate studies that would be prohibitive to perform on the entire set. Two uses of XRF WTA as a screening method are in progress. In the first, Gordon *et al.* (1982) obtained samples from the RAPS set that were identified by WTA as containing heavy loadings of particles from specific sources and analyzed them by instrumental neutron activation analysis (INAA) to obtain data for more elements than are observed by XRF. Second, Jaklevic (unpublished results) is subjecting some of the identified samples to XRD in an attempt to observe specific minerals. Many other sophisticated techniques (e.g. microscopic methods) could be applied to the identified samples.

The WTA method depends on the availability of a very large data set; a minimum of several hundred samples collected for ⩽ 12-h periods by each of three or more strategically located dichotomous samplers and analyzed by XRF. At least one recording wind vane should be included in the network. If Level II ambient sampling (see below) is done routinely, several months of data would be sufficient for application of WTA. In future sampling programs, receptors should be carefully placed with respect to sources and prevailing winds to maximize the potential for subsequent WTA. Appropriate wind data must also be obtained to define the influence of local vs regional sources.

AMBIENT SAMPLING AND ANALYSIS

No single protocol for sampling and analysis is applicable for all receptor model studies. Below we present an ideal hierarchy of sampling and analysis procedures of increasing sophistication and cost. However, if resources are limited, a decision must be made regarding the optimal division of those resources between source measurements, ambient sampling and analysis, meteorological measurements and interpretations. The division of effort among these categories to achieve a given level of reliability of final results (e.g. accuracy of the estimates of total suspended particulate contributions in two size fractions from various source classes) will vary for different areas. If most important sources in an area are common to many areas (e.g. coal- and oil-fired power plants, motor vehicles, marine aerosol, soil and regional sulfate), little additional study of sources may be needed. (However, the EPA 1980 Houston study demonstrated the need for at least some analyses of sieved, resuspended soil, whose composition may vary considerably from one area to another.) In these cases, most resources would be best devoted to ambient measurements. On the other hand, if an area contains sources that are unique or whose emissions are poorly described, considerable effort should be made to characterize those sources. For example, Houston has a number of industrial sources, especially in the petroleum and petrochemical industries, whose emissions are not well characterized. This fact, coupled with the unusual composition of Houston soil (e.g. its very large calcium concentration from marine shells), caused considerable confusion in the application of CMB interpretations. The final product would have been more reliable if more source information had been available.

Although ambient measurements are the focus of this section, ambient and source measurements should be considered together. Insofar as practical, similar collection and analysis methods should be used for both types of measurements (e.g. inlet cut point, type of sampler and filter media, analytical techniques).

Sampling

Particle samplers. In view of the availability and detailed characterization of the dichotomous sampler of the virtual impactor type, this device should be considered the basic sampler for receptor model applications and should be used at each sampling site. Particles of $> 2.5\,\mu m$ in diameter are chemically quite different from those of $< 2.5\,\mu m$. These groups should be kept separate to avoid the loss of information that occurs if they are mixed. Cascade impactors can separate particles into six or more fractions, but the additional fractionation often does not add much useful information. Studies on power plants and other combustion sources indicate that particles of $< 0.2\,\mu m$ in diameter constitute another qualitatively different group (Ondov *et al.*, 1981), but as yet no collector is routinely available. This group might be useful for identifying sources if enough of the particles survive coagulation to be observed. If, as appears probable, EPA adopts a standard for particles of $< 10\,\mu m$ in diameter, the upper cut point should probably be adjusted to that size. The 2.5-μm cut point, however, should be retained for identification purposes, whether or not a standard is established for the fine particles.

The filters routinely used in dichotomous samplers should have low blank values for elements of interest. Teflon and low blank Nuclepore filters meet this criterion for analyses by XRF. If it can be done without contaminating samples, the coarse-fraction filter should be coated with a substance that will prevent particle loss during handling. Either of the commercial samplers (Beckman and Sierra) with flow rates of $17\,\ell\,min^{-1}$ or the Berkeley sampler (Loo *et al.*, 1978) with a flow rate of $50\,\ell\,min^{-1}$ will provide enough sample for XRF and nuclear analysis. Mass can be determined by β-gauge (Jaklevic *et al.*, 1981; Courtney *et al.*, 1982).

The measurement of particulate mass and elemental concentrations represents the minimum level of ambient sampling and analysis for receptor modeling. Preferably, several additional measurements should be performed: particulate ionic species, carbon, organic compounds, crystalline substances (by XRD), individual particles (by electron microscopy), and gases. The additional measurements require that other samplers be run in parallel with the main samplers (namely, dichotomous samplers with either quartz filters for carbon and ionic analyses or Nuclepore filters for microscopy). XRD measurements can be performed on samples acquired with ordinary dichotomous samplers using Teflon membrane filters. However, optimum detectabilities for minor components require loadings of $> 100\,\mu g\,cm^{-2}$. Most crystalline species can be measured with minimum detectabilities of better than $5\,\mu g\,cm^{-2}$ using the Berkeley diffraction system (Thompson *et al.*, 1982). If costs are too high to permit the various samplers to be run concurrently, some savings are possible by collecting samples for microscopy (which does not require high loadings) for short periods in between collections for XRF, using the same samplers.

Sampling and analysis methods for organic compounds are still not as well developed as those for inorganic species, and there are no routine sample collection protocols. However, most analytical methods require large amounts of particulate material (i.e. 10's to 100's of milligrams). At present, the samples can be collected with a quartz hi-vol filter, although information on particle size will be sacrificed and the filters may adsorb some organic vapors. Size information could be obtained, if necessary, with and without a cyclone in front of the filter. Preferably, particles for organic analysis should be separated into two size groups, as the large-particle fraction may contain carbon in the form of bugs, pollen, carbonate, etc. whereas the compounds of interest for receptor modeling will be present mainly in the fine fraction. A developmental high-volume dichotomous sampler ($\sim 500\,\ell\,min^{-1}$) being tested by Solomon *et al.* (1983) may be of great value for collecting large masses of material.

Other samplers and instruments. The value of the data will be greatly enhanced if they include information on pollutant gas concentrations, condensation nuclei concentrations (CNC), and the light scattering coefficient (b_{scat}). These quantities can be measured in real time with rather high reliability by use of commercial instruments (if carefully calibrated and frequently serviced). Flame photometric total sulfur analyzers mainly measure SO_2. Chemiluminescence detectors are available for O_3 and for NO and, in a separate mode, for NO plus species that can be reduced to NO (mainly NO_2). The best available instruments, with the possible exception of the NO/NO_x detector, have sufficiently low detection limits to usually provide valid readings in an urban setting. Carbon monoxide can be measured accurately with infrared instruments. An integrating nephelometer is commonly used to measure light scattering, as is a condensation nuclei counter for CNC. As these are real-time devices, their readings should be frequently recorded by a data-logging device whose record can be read by a computer so that appropriate averages and standard deviations can be calculated. Depending on the number of input channels of the data-logger, it may also be used to store meteorological data.

In addition to the standard pollutant gases (CO, SO_2, NO/NO_x, and O_3), many other gases and vapors, if measured, could provide information of value for receptor modeling. Among these species are a wide range of hydrocarbons and other organic compounds and volatile forms of such elements as Hg, Se and the halogens. Methods for collection and measurement of some of these are reasonably well established (e.g. the collection of hydrocarbons in Teflon bags followed by analysis using gas chromatography, possibly coupled with mass spectrometry). Multiple columns in series have been used by Braman and Johnson (1974) to collect four different forms of airborne Hg, and Germani (1980) has used activated charcoal traps to collect volatile forms of several elements. However, despite the potential values of these types of measurements, we do not recommend their use in any but the most advanced, developmental receptor model field studies. Many of the collection and analysis methods are far from standard, and a more serious problem is that even if emission patterns are known, we know little about the modifications of these species during transit to the receptor. Modifications could result from reactions with species such as ·OH radicals, preferential condensation on airborne particles, or ground-level deposition.

A summary of a complete sampling station meeting all of the above needs is given in Table 1.

Sampling locations and periods. Most studies in urban areas will require at least six sampling stations (more if there are unique sources). All should be equipped with dichotomous samplers to permit subsequent XRF analyses, and at least one should be equipped with the samplers needed for the additional analyses discussed above and below. Before locations are picked, wind direction frequency distributions (wind 'roses') should be used to place stations so that they will receive emissions from major sources under predominant wind conditions, and so that the sources can be discriminated from each other by wind trajectories at clearly different angles from certain stations.

Because of the large meteorological changes that often occur between night and day, sample collection times should be ≤ 12 h in urban areas, and filters should be changed in the early morning and evening. If affordable, samples should be collected every 6 h, which greatly increases the probability of constant wind direction during sampling and, hence, the possibility for obtaining directional associations of collected particles. Reductions to less than 6 h are probably not cost-effective except to obtain the lightly loaded samples needed for microscopic studies, and even 6-h samples may not contain sufficient material for XRD.

Meteorology. A severe limitation of most previous studies has been the lack of adequate meteorological data. Hourly observations of instantaneous wind direction from the nearest airport are insufficient to determine the directions of important sources with respect to samplers. At least one sampling location should be equipped with a meteorological tower to obtain wind direction and speed with sufficient frequency to calculate hourly average values and standard deviations. If feasible, especially in a large study area, more than one station should be so equipped.

To relate results to dispersion models, one needs information on mixing heights and stability class. Classical methods such as balloons and radiosondes are helpful, but usually too expensive to use more than twice a day. Devices that yield more frequent readings should be considered: for example, a meteorological tower for short-term wind fluctuations or vertical temperature gradients ($\Delta T/\Delta z$) to calculate stability class, or an acoustic sounder to obtain mixing height. As development and testing increase their reliability and reduce their costs, various remote sensing devices should be considered for these applications. For example, laser radar (Lidar) instruments measure particle light-scattering vs altitude from ground level, or temperature versus altitude via population of rotational states of molecules of H_2O, CO_2, etc. These devices may be costly, but they permit the investigator to make frequent observations of temperature gradients, mixing heights, etc. at a fraction of the operating costs of traditional methods.

Analysis

As with the source studies and sampling, here we present a hierarchy of measurements of increasing sophistication and cost. The choice of analytical measurements should be based on the resources available for the study, the species to be measured, and the kinds of ambient samples collected. Obviously, the sampling methods are closely coupled with the analytical measurements; for example, certain analytical

Table 1. Suggested instruments for an ambient data collection station

Particulate sampler	Flow rate (l min^{-1})	Filter medium	Species measured	Analytical method(s)*
Dichotomous	17	Teflon	Mass	β-gauge, gravimetry
			Elements	XRF, INAA
		Quartz	Carbon	Combustion
			Ionic species (SO_4^{2-}, NO_3^-, etc.)	IC/chemistry
		Nuclepore	Individual particles	SEM/XRF
LBL (Berkeley) dichotomous	50	Teflon	Crystals, minerals	XRD
Hi-vol (possibly with cyclone or size-selective sampling head)	1000	Quartz	Organic compounds	GC/MS, HPLC, etc.

Gas sampler	Species measured	Output
Total sulfur	SO_2 and other S-containing gases	Real-time, need data-logging system
NO_x	NO, NO_2	Real-time, need data-logging system
O_3	O_3	Real-time, need data-logging system
CO	CO	Real-time, need data-logging system
Polyurethane or other column material	Organic vapors	Later extraction and analysis in laboratory

Other instrument	Parameter(s) measured	Output
Integrating nephelometer	Light scattering (b_{scat})	Real-time, need data-logging system
Condensation nuclei counter	Total particle number concentration (good indicator of fresh combustion aerosol)	Real-time, need data-logging system
Meteorological tower	Wind speed and direction	Real-time, need data-logging system (should record data at least once per minute)

* Definitions of abbreviations: GC/MS, gas chromatography/mass spectrometry; HPLC, high-performance liquid chromatography; IC, ion chromatography; INAA, instrumental neutron activation analysis; XRD, X-ray powder diffraction; XRF, X-ray fluorescence; SEM, scanning electron microscopy.

measurements cannot be performed unless the samples have been collected in a specific way.

Certain of the sample treatment and analytical methods discussed below remain under development. Some of the problems that require further research have been discussed in more detail by Hauser *et al.* (1983).

Particle analyses. The following criteria are considered in approximate order of increasing cost and information content.

Mass. At a minimum, the mass concentration in each size fraction should be measured. Historically, air filters were weighed under controlled humidity before and after particle collection. Mass concentrations for low-volume samples are now often measured by β-gauge, that is, by placing the filter and particle deposit between a radioactive β-emitting source and a detector and observing the reduction of count rate (Jaklevic *et al.*, 1981; Courtney *et al.*, 1982). This technique can be automated for measurement of large numbers of samples, and the results obtained for Teflon filters agree with those obtained by electrobalance (Courtney *et al.*, 1982).

Elements. Most receptor model studies involve such a large number of samples that traditional chemical methods are impractical. Automated XRF methods have been developed that involve the sequential use of three secondary fluorescers to emphasize elements in various regions of atomic number. They are capable of analyzing for up to 25 elements per sample over a few minutes (Stevens *et al.*, 1978; Loo *et al.*, 1978). Figure 1 compares typical concentrations of elements in urban atmospheres with detection limits for XRF and INAA (Dzubay, 1977; Cooper, 1973). Table 2 lists the elements that were easily, marginally, or not observable in the St. Louis RAPS.

XRF is fast and inexpensive, but cannot be used to determine accurately elements lighter than Al, nor routinely to observe various minute trace elements such as Se, As and Sb, which may yield more definitive signals of certain sources than more abundant elements. Thus, funds permitting, some samples should be analyzed using a different technique to extend the number of elements observed. A particularly useful combination is XRF followed by INAA. The latter technique is totally instrumental and, depending on mass loadings and filter blanks, can observe up to 40 elements (Kowalczyk *et al.*, 1982). Gordon *et al.* (1982) demonstrated the value of applying INAA to samples from the RAPS that had previously been subjected to XRF. Table 2 is based on their results, which show that the combination of XRF and INAA yielded data for 29 elements in the fine fraction and 32 in the coarse fraction for virtually every sampling period. An additional 18 elements in the fine fraction and 14 in the coarse fraction were observable in about half of the samples. Furthermore, 11 elements (Al, Ca, K, Cl, Ti, V, Mn, Fe, Zn, Br and Ba) can usually be measured by both methods to serve as a quality assurance check. Thorough INAA requires access to a reactor with a flux of at least $10^{13}\,\mathrm{n\,cm^{-2}\,s^{-1}}$ and a high-resolution Ge(Li) or intrinsic Ge γ-ray spectrometry system. The well thermalized neutrons of heavy water moderated reactors avoid inter-element interferences among light elements caused by fast neutrons in light water reactors. However, the fast neutrons allow one to observe Ni via an (n, p) reaction in light water reactors. Irradiations of two different lengths are needed for species of various half-lives (see below).

As presently performed, INAA requires considerably more sample handling and interpretation of results than XRF and, consequently, is considerably more costly per sample. Thus, XRF is the preferred screening technique. All samples could be subjected to XRF and the results used to identify only the most interesting samples for further analysis by INAA (for example, those heavily influenced by particles from certain major sources).

Another way to economize INAA is to use an approach taken by the Oregon Graduate Center (Watson, 1979; Core *et al.*, 1981). Complete INAA requires at least two irradiations of samples, the first for a few minutes to observe species with half-lives of ≤ 15 h, and the second for several hours to observe species with half-lives of up to several years. The former are much cheaper to perform, as the γ-ray spectra are simple enough to be resolved primarily by computer. Furthermore, in the former, results are usually available within a day after the irradiation, compared with about three weeks for long irradiations. The Oregon Graduate Center group has employed INAA cost-effectively by performing only the short irradiation following XRF analyses. Depending on the conditions, most of the following elements are usually observed: Na, Mg, Al, S, Cl, K, Ca, Ti, V, Mn, Cu, Br, La, Sm, Dy and W. However, some important elements that are normally observed in long irradiations (Cr, Co, As, Se and Sb) are usually sacrificed in this approach.

Many other techniques are used to observe elements in atmospheric particulate matter, including atomic absorption spectrophotometry (AAS), inductively coupled plasma atomic emission spectrometry (ICPS), spark-source mass spectrometry, various electrochemical methods (such as anodic stripping voltammetry), and chemical separations followed by colorimetry. Most require considerable labor per sample to obtain

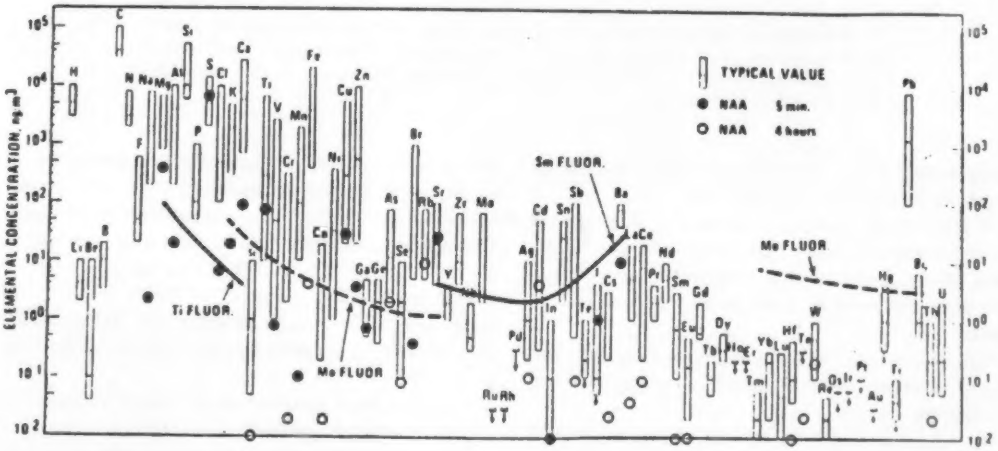

Fig. 1. Detection limits for XRF and INAA compared with the typical concentration values and ranges of elements in urban air.

Table 2. Detectability by XRF and INAA of elements in ambient samples of St. Louis

Element	Fine fraction XRF	Fine fraction INAA	Coarse fraction XRF	Coarse fraction INAA
Na	Not observable	Strong	Not observable	Strong
Mg	Not observable	Marginal	Not observable	Marginal
Al	Marginal	Strong	Strong	Strong
Si	Strong	Not observable	Strong	Not observable
P	Marginal	Not observable	Marginal	Not observable
S	Strong	Marginal	Strong	Not observable
Cl*	Strong	Strong	Strong	Strong
K	Strong	Strong	Strong	Strong
Ca	Strong	Marginal	Strong	Strong
Sc	Not observable	Strong	Not observable	Strong
Ti	Marginal	Marginal	Strong	Strong
V	Marginal	Strong	Marginal	Strong
Cr*	Marginal	Strong	Marginal	Strong
Mn	Marginal	Strong	Strong	Strong
Fe	Strong	Strong	Strong	Strong
Co	Not observable	Strong	Not observable	Strong
Ni	Marginal	Not observable	Marginal	Not observable
Cu	Marginal	Marginal	Marginal	Marginal
Zn	Strong	Strong	Strong	Strong
Ga	Marginal	Marginal	Marginal	Marginal
As	Marginal	Strong	Marginal	Strong
Se	Marginal	Strong	Marginal	Strong
Br	Strong	Strong	Strong	Strong
Rb	Marginal	Strong	Marginal	Strong
Sr	Marginal	Marginal	Marginal	Marginal
Mo	Not observable	Marginal	Not observable	Marginal
Ag	Not observable	Marginal	Not observable	Marginal
Cd	Marginal	Strong	Marginal	Marginal
In	Not observable	Marginal	Not observable	Marginal
Sn	Marginal	Not observable	Marginal	Not observable
Sb	Not observable	Strong	Marginal	Strong
I	Not observable	Strong	Not observable	Marginal
Cs	Not observable	Strong	Not observable	Strong
Ba	Marginal	Strong	Strong	Strong
La	Not observable	Strong	Not observable	Strong
Ce	Not observable	Strong	Not observable	Strong
Sm	Not observable	Strong	Not observable	Strong
Tb	Not observable	Marginal	Not observable	Marginal
Dy	Not observable	Marginal	Not observable	Strong
Yb	Not observable	Marginal	Not observable	Marginal
Lu	Not observable	Marginal	Not observable	Strong
Hf	Not observable	Strong	Not observable	Strong
Ta	Not observable	Marginal	Not observable	Strong
W	Not observable	Marginal	Not observable	Strong
Pb	Strong	Not observable	Strong	Not observable
Th	Not observable	Strong	Not observable	Strong

* May be unobservable because of filter blank.

accurate results, or (unlike XRF and INAA) require dissolution of samples prior to analysis. Moreover, any chemical manipulation greatly increases the chance of contamination by trace elements and loss of volatile or insoluble species.

Ionic species. We should not only know the concentrations of a large number of elements, but also the ionic forms of certain major species: for example, H^+, NO_3^-, NH_4^+ and SO_4^{2-}. These species can be measured on Teflon filters from dichotomous samplers, as in the Houston study (Johnson *et al.*, 1984; Dzubay *et al.*, 1982). In that case, the samples were preserved immediately after sampling by placing them in an extraction solution of NH_4Cl and $HClO_4$ of known concentrations. Total acid was determined by Grans titration, and NO_3^- and SO_4^{2-} were determined by ion chromatography (IC). Under favourable conditions, many anionic species can be determined by IC, including NO_2^- NO_3^-, SO_3^{2-}, SO_4^{2-}, Br^-, Cl^-, F^- and PO_4^{3-}, but the validity of NO_2^- and SO_3^{2-} is questionable because of possible instability of the oxidation states. Although NO_3^- can be measured accurately, there is considerable question about artifact formation (positive or negative) during collection. The filters recommended here, however, should have negligible artifacts for SO_4^{2-}. In principle, a number of cationic species (NH_4^+, Na^+, K^+, etc.) are detectable by IC using a different column, but this is rarely done, except for NH_4^+. The last may be observed by a colorimetric procedure (Rassum and Villarruz, 1963), as in the Houston study. Similarly, NO_3^- can be observed by a colorimetric method (Jassens and Dams, 1974), and water-soluble SO_4^{2-} can be assayed by placing a water extract of the filter in a filament and flash-vaporizing it into a total sulfur analyzer (Husar *et al.*, 1975).

Carbon. Carbon may account for 10–30% of the particulate mass, of which perhaps two thirds occurs in the fine fraction. At a minimum, the total carbon should be measured, possibly by combustion and measurement of the CO_2 released. The carbon released as a function of temperature in differential thermal analysis also should be measured. Organic materials are released at fairly low temperatures and elemental carbon at much higher temperatures (Johnson *et al.*, 1981). If the latter is coupled with measurement of light transmission through the filter versus temperature, additional data are provided on the absorption coefficient of particulate matter, b_{ap}, a factor that contributes appreciably to visibility degradation (about 6–13% in Houston) (Dzubay *et al.*, 1982). In the past, the fraction of carbon extracted by organic solvents such as benzene was frequently measured. Grosjean (1975) showed, however, that this nonpolar solvent does not give complete recovery of particulate organic carbon, and recommended that a nonpolar and a polar solvent be used as a binary mixture or in sequence.

Converting carbon to CO_2 and then using a low-background β proportional counter to measure the ^{14}C-to-total-carbon ratio has been used to considerable advantage in a few studies (Cooper *et al.*, 1981). This ratio indicates the fraction of particulate carbon arising from fossil fuels vs that from 'modern' carbon sources, such as the combustion of vegetation or recently living wood. Fossil fuels contain no ^{14}C, whereas modern sources have a ^{14}C-to-total-carbon ratio approximately equal to that of the atmosphere. In the Houston study, even using the very-low-background counter at the National Bureau of Standards, the sample mass requirements necessitated use of a quartz hi-vol filter behind an inlet consisting of six coaxial cylinders that removed most particles of $> 4\ \mu m$ in diameter. Recently, considerable progress has been made in increasing the sensitivity of ^{14}C measurements by use of nuclear particle accelerators. With this method, samples containing $< 50\ \mu g$ carbon (vs milligram quantities with counters) can be used, thus eliminating the need for special high-volume collections (Currie and Klouda, 1982). At present, few U.S. laboratories can perform the ^{14}C measurements.

Organic compounds. Although not used in the Houston study, one of the potentially most useful techniques for receptor model applications is the analysis of particles (and, possibly, vapors) for organic compounds. Such a multitude of organic compounds is present in the atmosphere (Daisey, 1980; Daisey and Kneip, 1981; Simoneit and Mazurek, 1981) that measurement of many members of appropriate compound classes could provide very high resolution of sources, if certain conditions were met. The study of organic compounds is needed particularly for the identification and assay of particles from sources that release mostly carbonaceous particles that provide little or no signal via the elements observed by XRF and INAA (e.g. engines burning diesel fuel or unleaded gasoline, home heating, refineries, coke ovens, petrochemical plants). Perhaps because of the carcinogenic nature of many of its members, the series of polynuclear aromatic hydrocarbons (PAHs) probably has been the most extensively studied (Daisey, 1980; Daisey *et al.*, 1979; Hites, 1981; Hites *et al.*, 1977; Fox and Staley, 1976). The PAHs do fulfill a major need of markers for receptor models in that relative intensities of the various members appear to differ depending on the type of source, the combustion temperature, etc. However, other classes of organics, especially the long-chain aliphatic compounds, contribute much greater masses of material (Daisey, 1980; Lamb *et al.*, 1980; Simoneit and Mazurek, 1981). Thus, the study of organic compounds should include observations of several classes of compounds. Before organic compounds can be routinely used, the following investigations should be made:

(a) Carefully review existing source emissions data to identify compound classes that may be of potential use in distinguishing source types. Then perform extensive studies on major sources of carbonaceous particles to determine the compound classes that show the greatest pattern differences between source types.

(b) For those classes that appear promising, establish the fraction of each member that is typically in the gas phase, and determine how that fraction changes with distance from the source or variation of ambient conditions. For these studies, both filters and downstream polyurethane plugs should be analyzed (Yamasaki *et al.*, 1982; Glotfelty *et al.*, 1983). Tenax plugs have commonly been used for this purpose, but flow rates through them are so low that it is difficult to collect enough material for some analyses. If polyurethane is used, a series of columns should first be used to ensure that the species of interest do not penetrate the column over the sampling duration.

(c) Identify the organic compounds that, upon release, are stable or react slowly enough that transformation can be ignored.

Analytical methods for PAHS (e.g. GC and high-performance liquid chromatography) are well developed. A calibration standard (PAH in acetonitrile) and a particulate matter Standard Reference Material certified for certain PAH compounds have recently become available through the National Bureau of Standards (May *et al.*, 1982). Capillary column GC/MS has proven to be a valuable analytical method for multi-compound analysis of nonpolar, volatile organic compounds in particulate matter. The accuracy of this method, however, is limited by the availability of standard compounds. Methods for more polar organic compounds and for labile compounds are not well developed.

Crystals. A prospective method to identify particle sources is to observe various crystalline materials according to their XRD patterns. Thompson *et al.* (1982) at Lawrence Berkeley Laboratory (LBL) have developed an XRD unit specifically for analysis of air filter samples. With it they are able to observe a number of minerals in particulate material, including quartz, calcite, dolomite, halite, mascagnite and lead ammonium sulfate. The ability to observe minerals in the Houston study was limited by the particulate loading of the dichotomous sampler filters (typically 70 $\mu g\,cm^{-2}$) and the fact that the LBL instrument was not yet fully operational. Particulate loadings of $> 200\ \mu g\,cm^{-2}$ are optimum for XRD, and the sensitivity of the LBL instrument has now been improved to give detectabilities of $< 5\ \mu g\,cm^{-2}$ for major crystalline species. If greater sensitivity is required, the XRD technique should be applied to samples collected with hi-vol filters with and without cyclones upstream.

Individual particle analysis. In the Houston study, Johnson *et al.* (1984) used computer-assisted scanning electron microscopy with automated image analysis and X-ray energy spectroscopy (SAX) to scan the filters and classify individual particles according to their shapes, sizes, and XRF analyses for several elements. Each particle was classified according to source type by comparing the measured parameters against those of 25 particle classes stored in a library. Most of the library classes were developed from previous studies in other areas, although Houston soils were observed and included. The coarse-fraction samples from Houston had loadings of 50–60 $\mu g\,cm^{-2}$, which provided a particle spacing on the filters that was appropriate for observation. Particles could be sorted into about a dozen categories. Analyses of the fine-fraction samples were much less successful, because the much higher particle number density (especially of secondary sulfate particles, which coated most of the area) made it difficult to observe other types of particles.

This method is still highly experimental but holds great promise because of its high resolution, that is, its ability to sort particles into as many as 25 classes. In future receptor model studies, particulate samples from the major sources should be collected with loadings appropriate for SAX to build a library of characteristics of particles from sources in the area. In both source and ambient sampling, fine-fraction samples for SAX should be collected for shorter times than the coarse fractions, to obtain optimal spacing of particles on the filters.

Gases. Very thorough studies, such as the RAPS, have included measurement of several gases (e.g. SO_2, O_3, NO, NO_2, CO) simultaneously with particle collection. However, most receptor model treatments have focused on the composition of the particles and have largely ignored the gas data. But the gases also carry information about sources (e.g. CO from motor vehicles, SO_2 from power plants and smelters, NO from high-temperature processes and consequent deficiency of O_3 from reaction with NO). The gases should be measured at both sources and receptors, and the information used along with the other parameters in receptor model calculations. Gas measurements are real-time measurements with < 1-min response times, so they can add much information about short-term fluctuations related to rapid changes of meteorological conditions (especially wind direction). Gas data also permit the modeler to draw connections between receptor model results and those of more traditional dispersion modeling.

Levels of sampling and analysis

From the above discussion it is clear that a wide variety of ambient sampling and analysis methods are available. Below, we recommend three levels of increasing sophistication and cost that parallel the levels given above for determination of compositions of source materials. We assume that most state and local agencies are now equipped with fairly standard monitoring equipment typical of the National Air Surveillance Network (NASN). Although more sophisticated equipment is desirable, some receptor model work is possible with careful use of this equipment. Thus, an agency can start with such equipment and do receptor model studies at fairly reasonable cost and with only minor changes of procedures and materials. As funds for capital equipment become available, the system can be upgraded to Levels II and III.

Level I. This level would entail operation of three or more hi-vol samplers for ≤ 24-h periods. Preferably, these samplers would be equipped with modern inlet heads and filters with low blank values (i.e. not standard glass fiber filters). Samples would be analyzed for mass, SO_4^{2-}, and a number of metals by AAS, ICPS, or other chemical methods. Level I also would include standard measurement of the gases SO_2, CO, O_3, and NO/NO_x, and acquisition of meteorological data from the nearest airport or National Weather Service station.

Level II. In this level, three or more dichotomous samplers with Teflon filters would be operated for ≤ 12-h periods. All samples would be analyzed for mass by β-gauge or electrobalance and for elements by XRF. Samples obtained by dichotomous samplers with quartz filters would be analyzed for carbon and ionic species. Pollutant gases would be monitored by high-quality real-time instruments connected to a data-logging system. The network would include at least one meteorological station equipped with a wind vane connected to a data-logger.

Level III. This level would encompass all of the systems and measurements listed in Table 1. However, not all of these would need to be added at once. As funds become available, a local agency might wish to equip one key station with as many of the Level III instruments as possible, while maintaining the rest of the network at Level II. As funds permitted, it would be desirable to upgrade at least three stations to Level III to permit use of triangulation in WTA to identify important sources. Note that a number of the Level III methods, especially those involving minerals, individual particles, vapor-phase species, and organic compounds, are still under development and testing. We anticipate, however, that these methods will have been validated and standardized by the time any agency would be ready to move to complete Level III studies.

SUMMARY AND RECOMMENDATIONS

Receptor model studies of the sources of airborne particles in an area can be initiated at rather small cost with equipment that is much the same as is used in monitoring. The ability of state and local agencies to perform receptor model studies would be greatly enhanced if EPA would provide the best available data on compositions of particles from various types of ubiquitous sources (e.g. coal- and oil-fired power plants, motor vehicles, etc.), receptor model computer packages, and other assistance. Regarding the latter, a whole array of microscopic tools for observing the characteristics of individual particles are just beginning to be exploited. These tools include the analytical scanning electron microscope and the laser microprobe mass analyzer (LAMMA) (Wieser *et al.*, 1980). The use of ^{14}C, discussed above, is but one example of the measurement of naturally occurring isotopes. In the future, isotopes of many light elements, including B, C (including ^{13}C), N, O and S, may prove useful. Thus, although the equipment and methods for receptor model studies have improved enormously over the past few years, the field is still developing, and we can look forward to continued improvements that will enable even more definitive receptor model studies.

We cannot stress too emphatically the need to coordinate meteorological measurements with other field measurements. Receptor modelers, most of whom are basically atmospheric or analytical chemists, have tended to focus on the collection and analysis of chemical species. However, we can greatly enhance the value of data sets by collecting samples over periods sufficiently brief to avoid excessive wind variability and by simultaneously obtaining on-site meteorological data, especially detailed wind direction data.

Acknowledgements—The authors thank the EPA personnel, especially Robert K. Stevens, Thompson G. Pace and Thomas G. Dzubay, for their excellent services to the receptor modeling community, including design and conduct of the Houston study, provision to the investigators of data from that and the simulated study, and encouragement to all of us to interpret and publish results of this valuable exercise. We also thank our colleagues in the receptor model field for their contributions and suggestions regarding this paper.

REFERENCES

Alpert D. J. and Hopke P. K. (1980) A quantitative determination of the sources in the Boston urban aerosol. *Atmospheric Environment* **14**, 1137–1146.

Alpert D. J. and Hopke P. K. (1981) A determination of the sources of airborne particles collected during the Regional Air Pollution Study. *Atmospheric Environment* **15**, 675–687.

Armstrong J. A., Russell P. A., Sparks L. E. and Drehmel D. C. (1981) Tethered balloon sampling systems for monitoring air pollution. *J. Air Pollut. Control Ass.* **31**, 735–743.

Braman R. S. and Johnson D. L. (1974) Selective absorption tubes and emission technique for determination of ambient forms of mercury in air. *Envir. Sci. Technol.* **8**, 996–1003.

Cass G. R. and McRae G. J. (1981) Source–Receptor reconciliation of South Coast Air Basin particulate air quality data. Report for California Air Resources Board.

Chang S.-N., Hopke P. K., Rheingrover S. W. and Gordon G. E. (1982) Target-transformation factor analysis of wind-trajectory-selected samples. Paper 82-21.1, Air Pollution Control Association.

Cooper J. A. (1973) Urban aerosol trace element ranges and typical values. BNWL-SA-4690, Battelle Pacific Northwest Laboratories, Richland, Washington.

Cooper J. A., Currie L. A. and Klouda G. A. (1981) Assessment of contemporary carbon combustion source contributions to urban air particulate levels using carbon-14 measurements. *Envir. Sci. Technol.* **15**, 1045–1050.

Core J. E., Hanrahan P. L. and Cooper J. A. (1981) Air particulate control strategy development: a new approach using chemical mass balance methods. In *Atmospheric Aerosol: Source/Air Quality Relationships* (edited by Macias and Hopke), pp. 107–123. ACS Symp. Series No. 167, American Chemical Society, Washington, D.C.

Courtney W. J., Shaw R. W. and Dzubay T. G. (1982) Precision and accuracy of a beta-gauge for aerosol mass determinations. *Envir. Sci. Technol.* **16**, 236–239.

Currie L. A. and Klouda G. A. (1982) Nuclear and chemical dating techniques: interpreting the environmental record. In *Counters, Accelerators and Chemistry* (edited by Currie), Chapter 10. ACS Symp. Series No. 176, American Chemical Society, Washington, D.C.

Daisey J. M. (1980) Organic compounds in urban aerosols. *Ann. N. Y. Acad. Sci.* **338**, 50–69.

Daisey J. M. and Kneip T. J. (1981) Atmospheric particulate organic matter: Multivariate models for identifying sources and estimating their contributions to the ambient aerosol. In *Atmospheric Aerosol: Source/Air Quality Relationships* (edited by Macias and Hopke), pp. 197–221. ACS Symp. Series No. 167, American Chemical Society, Washington, D.C.

Daisey J. M., Leyko M. A. and Kneip T. J. (1979) Source identification and allocation of polynuclear aromatic hydrocarbon compounds in the New York City aerosol: Methods and applications. In *Polynuclear Aromatic Hydrocarbons* (edited by Jones and Leber), pp. 201–215. Ann Arbor Science Publishers, Ann Arbor, Michigan.

Dzubay T. G. (1977) Recent advances in energy-dispersive XRF analysis of aerosol samples. *Trans. Am. Nucl. Soc.* **27**, 203–204.

Dzubay T. G. (1980) Chemical element balance method applied to dichotomous sampler data. *Ann. N. Y. Acad. Sci.* **338**, 126–144.

Dzubay T. G., Stevens R. K., Lewis C. W., Hern D., Courtney W. J., Tesch J. W. and Mason M. A. (1982) Visibility and aerosol composition in Houston, Texas. *Envir. Sci. Technol.* **16**, 514–525.

Dzubay T. G., Stevens R. K., Balfour W. D., Williamson H. J., Cooper J. A., Core J. E., De Cesar R. T., Crutcher E. R., Dattner S. L., Davis B. L., Heisler S. L., Shah J. J., Hopke P. K. and Johnson D. L. (1984) Interlaboratory comparison of receptor model results for Houston aerosol. *Atmospheric Environment* **18**, 1555–1566.

Fox M. A. and Staley S. W. (1976) Determination of polycyclic aromatic hydrocarbons in atmospheric particulate matter by high pressure liquid chromatography coupled with fluorescence techniques. *Analyt. Chem.* **48**, 992–998.

Germani M. S. (1980) Selected studies of four high-temperature air-pollution sources. Ph.D. dissertation, Department of Chemistry, University of Maryland, College Park, Maryland.

Gladney E. S., Small J. A., Gordon G. E. and Zoller W. H. (1976) Composition and size distribution of in-stack particulate material at a coal-fired power plant. *Atmospheric Environment* **10**, 1071–1077.

Glotfelty D. E., Taylor A. W. and Zoller W. H. (1983) Atmospheric dispersion of vapors: are molecular properties unimportant? *Science, Wash.* **219**, 843–845.

Gordon G. E. (1980) Receptor models. *Envir. Sci. Technol.* **14**, 792–800.

Gordon G. E., Zoller W. H., Kowalczyk G. S. and Rheingrover S. W. (1981) Composition of source components needed for aerosol receptor models. In *Atmospheric Aerosol: Source/Air Quality Relationships* (edited by Macias and Hopke), pp. 51–74. ACS Symp. Series No. 167. American Chemical Society, Washington, D.C.

Gordon G. E., Rheingrover S. W. and Aras N. K. (1982) Determining compositions of particles from major sources from ambient samples. *Trans. Am. Nucl. Soc.* **41**, 199–200.

Greenberg R. R., Zoller W. H. and Gordon G. E. (1978a) Composition and size distributions of particles released in refuse incineration. *Envir. Sci. Technol.* **12**, 566–573.

Greenberg R. R., Gordon G. E., Zoller W. H., Jacko R. B., Neuendorf D. W. and Yost K. J. (1978b) Composition of particles emitted from the Nicosia municipal incinerator. *Envir. Sci. Technol.* **12**, 1329–1332.

Grosjean D. (1975) Solvent extraction and organic carbon determination in atmospheric particulate matter: the organic extraction-organic carbon analyzer (OE-OCA) technique. *Analyt. Chem.* **47**, 797–805.

Hauser T. R., Scott D. and Midgett M. R. (1983) Air monitoring: research needs. *Envir. Sci. Technol.* **17**, 86A–96A.

Heinsohn R. J., Davis J. W. and Knapp K. T. (1980) Dilution source sampling system. *Envir. Sci. Technol.* **14**, 1205–1209.

Henry R. C., Lewis C. W., Hopke P. K. and Williamson H. J. (1984) Review of receptor model fundamentals. *Atmospheric Environment* **18**, 1507–1515.

Hites R. A. (1981) Sources and fates of atmospheric polycyclic aromatic hydrocarbons. In *Atmospheric Aerosol: Source Air Quality Relationships* (edited by Macias and Hopke), pp. 187–196. ACS Symp. Series No. 167, American Chemical Society, Washington, D.C.

Hites R. A., LaFlamme R. E. and Farrington J. W. (1977) Sedimentary polycyclic aromatic hydrocarbons: the historical record. *Science, Wash.* **198**, 829–831.

Houck J. E., Core J. E. and Cooper J. A. (1982) *Source Sampling for Receptor Oriented Source Apportionment*. NEA, Inc., Beaverton, Oregon.

Husar J. D., Husar R. B. and Stubits P. K. (1975) Determination of submicrogram amounts of atmospheric particulate sulfate. *Analyt. Chem.* **47**, 2062–2065.

Jaklevic J. M., Gatti R. C., Goulding F. S. and Loo B. W. (1981) A beta-gauge method applied to aerosol samples. *Envir. Sci. Technol.* **15**, 680–686.

Jassens M. and Dams R. (1974) Determination of cadmium in air particulates by flameless atomic absorption spectrometry with a graphite tube. *Analytica chim. Acta* **70**, 25–31.

Johnson D. L., Davis B. L., Dzubay T. G., Hasan H., Crutcher E. R., Courtney W. J., Jaklevic J. M., Thompson A. C. and Hopke P. K. (1984) Chemical and physical analyses of Houston aerosol for interlaboratory comparison of source apportionment procedures. *Atmospheric Environment* (this issue).

Johnson R. L., Shah J. J., Cary R. A. and Huntzicker J. J. (1981) An automated thermal-optical method for the analysis of carbonaceous aerosol. In *Atmospheric Aerosol: Source/Air Quality Relationships* (edited by Macias and Hopke), pp. 223–234. ACS Symp. Series No. 167, American Chemical Society, Washington, D.C.

Klein D. H., Andren A. W., Carter J. A., Emery J. F., Feldman C., Fulkerson, W., Lyon W., Ogle J. C., Talmi Y., Van Hook R. I. and Bolton N. (1975) Pathways of thirty-seven trace elements through a coal-fired power plant. *Envir. Sci. Technol.* **9**, 973–979.

Kowalczyk G. S., Gordon G. E. and Rheingrover S. W. (1982) Identification of atmospheric particulate sources in Washington, D.C. using chemical element balances. *Envir. Sci. Technol.* **16**, 79–90.

Lamb S. I., Petrowski C., Kaplan I. R. and Simoneit B. R. T. (1980) Organic compounds in urban atmospheres: a review of distribution, collection and analysis. *J. Air Pollut. Control Ass.* **30**, 1098–1115.

Loo B. W., French W. R., Gatti R. C., Goulding F. S., Jaklevic J. M., Llacer J. and Thompson A. C. (1978) Large scale measurement of airborne particulate sulfur. *Atmospheric Environment* **12**, 759–771.

May W. E., Chesler S. N., Hertz H. S. and Wise S. A. (1982) Analytical standards and methods for the determination of polynuclear aromatic hydrocarbons in environmental samples. *Int. J. Envir. Analyt. Chem.* **12**, 259–275.

Ondov J. M., Ragaini R. C. and Biermann A. H. (1979a) Elemental emissions from a coal-fired power plant. Comparison of a Venturi wet scrubber system with a cold-side electrostatic precipitator. *Envir. Sci. Technol.* **13**, 598–607.

Ondov J. M., Ragaini R. C. and Biermann A. H. (1979b) Emissions and particle-size distributions of minor and trace elements at two western coal-fired power plants equipped with cold-side electrostatic precipitators. *Envir. Sci. Technol.* **13**, 946–953.

Ondov J. M., Biermann A. H., Heft R. E. and Koszykowski R. F. (1981) Elemental composition of atmospheric fine particles emitted from coal burned in a modern electric power plant equipped with a flue-gas desulfurization system. In *Atmospheric Aerosol: Source Air Quality Relationships* (edited by Macias and Hopke), pp. 173–186. ACS Symp. Series No. 167, American Chemical Society, Washington, D.C.

Ondov J. M., Zoller W. H. and Gordon G. E. (1982) Trace element emissions on aerosols from motor vehicles. *Envir. Sci. Technol.* **16**, 318–328.

Pierson W. R. and Brachaczek W. W. (1983) Particulate matter associated with vehicles on the road—II. *Aerosol Sci. Technol.* **2**, 1–40.

Pilat M. J., Ensor D. S. and Bosch J. C. (1970) Source test cascade impactor. *Atmospheric Environment* **4**, 671–679.

Rassum J. R. and Villarruz P. A. (1963) Determination of ammonia by the indophenol method. *J. Am. Water Works Assn.* **55**, 675–678.

Rheingrover S. W. and Gordon G. E. (1980) Identifying locations of dominant point sources of elements in urban atmospheres from large, multi-element data sets. In *Proc. Fourth Int. Conf. on Nuclear Methods in Environmental and Energy Research*, pp. 82–95. University of Missouri, Columbia, Missouri.

Rom J. J. (1972) Maintenance, calibration, and operation of isokinetic source sampling equipment. APTD-0576, U.S. Environmental Protection Agency.

Schrag M. P. and Johnson G. L. (1978) Fine particle emissions information system (FPEIS) users guide. EPA-600 8-78-006, U.S. Environmental Protection Agency.

Shacklette H. T., Hamilton J. C., Boerngen J. A. and Bowles J. M. (1971) *Elemental Composition of Surficial Materials in the Conterminous United States*. Professional Paper 574-D.

U.S. Geological Survey.

Simoneit B. R. T. and Mazurek M. A. (1981) Air pollution: the organic components. *CRC Crit. Rev. Envir. Control* **11**, 219–276.

Small M., Germani M. S., Small A. M., Zoller W. H. and Moyers J. L. (1981) Airborne plume study of emissions from the processing of copper ores in Southeastern Arizona. *Envir. Sci. Technol.* **15**, 293–299.

Solomon P. A., Moyers J. A. and Fletcher R. A. (1983) High-volume dichotomous virtual impactor for the fractionation and collection of particles according to aerodynamic size. *Aerosol Sci. Technol.* **2**, 455–464.

Stevens R. K. and Pace T. G. (1984) Overview of the Mathematical and Empirical Receptor Models Workshop (Quail Roost II). *Atmospheric Environment* **18**, 1499–1506.

Stevens R. K., Dzubay T. G., Russwurm G. and Rickel D. (1978) Sampling and analysis of atmospheric sulfates and related species. *Atmospheric Environment* **12**, 55–68.

Thompson A. C., Jaklevic J. M., O'Conner B. H. and Morris C. M. (1982) X-ray powder diffraction system for chemical speciation of particulate aerosol samples. *Nucl. Instrum. Methods Phys. Res.* **198**, 539–546.

Watson J. G. (1979) Chemical element balance receptor model methodology for assessing the sources of fine and total suspended particulate matter in Portland, Oregon. Ph.D. dissertation, Oregon Graduate Center, Beaverton, Oregon.

Wieser P., Wurster R. and Seiler H. (1980) Identification of airborne particles by laser induced mass spectroscopy. *Atmospheric Environment* **14**, 485–494.

Yamasaki H., Kuwata K. and Miyamoto H. (1982) Effects of ambient temperature on aspects of airborne polycyclic aromatic hydrocarbons. *Envir. Sci. Technol.* **16**, 189–194.